Cul. H02325. 65- 0160850

65

AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

R. BAER
UNIVERSITY OF ILLINOIS

S. EILENBERG
COLUMBIA UNIVERSITY

D. C. LEWIS, JR.
THE JOHNS HOPKINS UNIVERSITY

A. WINTNER
THE JOHNS HOPKINS UNIVERSITY

WITH THE COÖPERATION OF

L. V. AHLFORS
C. B. ALLENDOERFER
W. L. CHOW
P. R. HALMOS
P. HARTMAN
I. KAPLANSKY
W. S. MASSEY
H. SAMELSON
J. J. STOKER
O. SZÁSZ
R. M. THRALL
A. WEIL

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

510.5
C C 2

VOLUME LXXIII
1951

CALCUTTA UNIVERSITY
CENTRAL LIBRARY.

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

160,680

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

INDEX

PAGE

Denjoy, Arnaud. Une extension du théorème de Vitali, . . . 314

Boothby, William M. The topology of regular curve families with multiple saddle points, . . . 405

Boothby, William M. The topology of the level curves of harmonic functions with critical points, . . . 512

Borg, Göran. On the point spectra of $y'' + (\lambda - q(x))y = 0$, . 122

Brahana, H. R. Finite metabelian groups and the lines of a projective four-space, . . . 539

Brauer, Alfred. On the irreducibility of polynomials with large third coefficient. II, . . . 717

Cronin, Jane. A definition of degree for certain mappings in Hilbert space, . . . 765

Dieudonné, Jean. Linearly compact spaces and double vector spaces over sfields, . . . 13

Dieudonné, Jean. Orthogonal and unitary groups over the rational field, . . . 940

Dubisch, Roy and Sam Perlis. On total nilpotent algebras, . . . 439

Errata, . . . 963

Floyd, E. E. Some retraction properties of the orbit decomposition spaces of periodic maps, . . . 363

Garabedian, P. R. Asymptotic identities among periods of integrals of the first kind, . . . 107

Gilbarg, David. The existence and limit behavior of the one-dimensional shock layer, . . . 256

Haimo, Franklin. A representation for Boolean algebras, . . . 725

Hartman, Philip. The number of L^2-solutions of $x'' + q(t)x = 0$, 635

Hartman, Philip. On bounded Green's kernels for second order linear ordinary differential equations, . . . 646

Hartman, Philip. On geodesic coordinates, . . . 949

Hartman, Philip. On linear second order differential equations with small coefficients, . . . 955

Hartman, Philip. On the eigenvalues of differential equations, . 657

PAGE

HARTMAN, PHILIP and AUREL WINTNER. On the problems of geodesics in the small, . . . 132

HARTMAN, PHILIP and AUREL WINTNER. On the asymptotic curves of a surface, . . . 149

HARTMAN, PHILIP and AUREL WINTNER. On the classical transcendents of mathematical physics, . . . 381

HARTMAN, PHILIP and AUREL WINTNER. On the non-increasing solutions of $y'' = f(x, y, y')$, . . . 390

HARTMAN, PHILIP and AUREL WINTNER. Gaussian curvature and local embedding, . . . 876

HARTMAN, PHILIP and AUREL WINTNER. On an oscillation criterion of Liapounoff, . . . 885

HERSTEIN, I. N. A generalization of a theorem of Jacobson, . . . 756

HIRSCHMAN, I. I., JR. On approximation by non-dense sets of translates, . . . 773

HOSTINSKY, L. AILEEN. Direct decomposition in lattices, . . . 741

HSU, L. C. Generalized Stieltjes-Post inversion formula for integral transforms involving a parameter, . . . 199

HSU, L. C. On the asymptotic evaluation of a class of multiple integrals involving a parameter, . . . 625

JENKINS, JAMES A. On an inequality of Golusin, . . . 181

JENKINS, JAMES A. Generalization of a theorem of Mandelbrojt, . . . 807

JÓNSSON, BJARNI and ALFRED TARSKI. Boolean algebras with operators, 891

KAPLANSKY, IRVING. Locally compact rings. II, . . . 20

KODAIRA, KUNIHIKO. The theorem of Riemann-Roch on compact surfaces, . . . 813

LANDWEBER, L. An iteration formula for Fredholm integral equations of the first kind, . . . 615

LEAVITT, WILLIAM G. On systems of linear differential equations, . 690

LEVITZKI, JAKOB. Prime ideals and the lower radical, . . . 25

LEWIS, D. C. Differential equations referred to a variable metric, . 48

MACKEY, GEORGE W. On induced representations of groups, . . . 576

MILLS, W. H. The m-th power residue symbol, . . . 59

MILLS, W. H. Reciprocity in algebraic number fields, . . . 65

MITCHELL, JOSEPHINE. On the spherical convergence of multiple Fourier series, . . . 211

PAGE

MORDELL, L. J. The reciprocity formula for Dedekind sums, . . 593

NAGUMO, MITIO. A theory of degree of mapping based on infinitesimal analysis, ... 485

NAGUMO, MITIO. Degree of mapping in convex linear topological spaces, ... 497

NAKAYAMA, TADASI. Generalized Galois theory for rings with minimum condition, ... 1

NEHARI, ZEEV. Extremal problems in the theory of bounded analytic functions, ... 78

NEWMAN, D. J. The evaluation of the constant in the formula for the number of partitions of n, ... 599

PAN, T. K. Hypergeodesics and dual hypergeodesics, ... 556

PERLIS, SAM. See Dubisch and Perlis, page 439.

PETTIS, B. J. Uniform Cauchy points and points of equicontinuity, 602

PUTNAM, C. R. On commutators of bounded matrices, ... 127

PUTNAM, C. R. On normal operators in Hilbert space, ... 357

PUTNAM, CALVIN R. and AUREL WINTNER. Linear differential equations with almost periodic or Laplace transform coefficients, 792

RAMANATHAN, K. G. The theory of units of quadratic and hermitian forms, ... 233

RICKART, C. E. Isomorphic groups of linear transformations, II, . 697

SCHENKMAN, EUGENE. A theory of subinvariant Lie algebras, . . 453

SEGAL, I. E. Equivalences of measure spaces, ... 275

SENIOR, JAMES K. Partitions and their representative graphs, . . 663

SHERMAN, S. Order in operator algebras, ... 227

STOLL, R. R. Homomorphisms of a semigroup onto a group, . . 475

SZÁSZ, OTTO. Tauberian theorems for summability (R_1), ... 779

TARSKI, ALFRED. See Jónsson and Tarski, page 891.

THIELMAN, H. P. A note on a functional equation, ... 482

THRALL, R. M. and G. de B. ROBINSON. Supplement to a paper of G. de B. Robinson, ... 721

TRUESDELL, C. A form of Green's transformation, ... 43

WINTNER, AUREL. See Hartman and Wintner, page 132.

WINTNER, AUREL. See Hartman and Wintner, page 149.

WINTNER, AUREL. See Hartman and Wintner, page 173.

PAGE

WINTNER, AUREL. See Hartman and Wintner, page 381.

WINTNER, AUREL. See Hartman and Wintner, page 390.

WINTNER, AUREL. See Hartman and Wintner, page 876.

WINTNER, AUREL. See Hartman and Wintner, page 885.

WINTNER, AUREL. See Putnam and Wintner, page 792.

WINTNER, AUREL. On the small divisors in integrations by Laplace transforms, ... 173

WINTNER, AUREL. On the non-existence of conjugate points, . . 368

WINTNER, AUREL. On Riemann metrics of constant curvature, . . 569

YOOD, BERTRAM. Banach algebras of continuous functions, . . . 30

ZAREMBA, S. K. On nets of curves on non-orientable surfaces, . . 186

510.5
.002

GENERALIZED GALOIS THEORY FOR RINGS WITH MINIMUM CONDITION.*

By Tadasi Nakayama.

In a previous note [1] the writer established a Galois theory for general rings with minimum condition which specializes in the case of division rings to Jacobson's Galois theory with outer groups of automorphisms.[2] On the other hand, Jacobson and Cartan have, independently, given a generalized Galois theory for division rings which embraces the theory of that type and the theory of commuter systems.[3] The present work is to propose a similar generalization for rings with minimum condition, or primary rings, to be exact.[4] This we shall do by combining the results in the above paper with those in another note,[5] in which the writer studied commuter systems in general rings with minimum condition.

0. Let R be,[6] throughout in this note, a ring with unit element 1 and satisfying the minimum condition (whence the maximum condition) for ideals. Let $\mathfrak{A}$ be the (absolute) endomorphism ring of the module R. The left and right multiplication rings R_l, R_r are subrings of $\mathfrak{A}$. Let $\mathfrak{B}$ be a subring of $\mathfrak{A}$ that contains R_r; $\mathfrak{B} \supseteq R_r$. Its commuter system $V_{\mathfrak{A}}(\mathfrak{B})$ in $\mathfrak{A}$ is a subring of $V_{\mathfrak{A}}(R_r) = R_l$, and has a form $V_{\mathfrak{A}}(\mathfrak{B}) = S_l$, the ring of left multiplications on R of S, with a subring S of R.

Suppose that R is (right-) *regular* with respect to $\mathfrak{B}$, in the sense of [8] or [9].[7] Then R is regular with respect to $S_l = V_{\mathfrak{A}}(\mathfrak{B})$ too, and

---

* Received March 8, 1949; revised June 14, 1950. The writer is very grateful to the referee, whose kind comments were very useful, as well as valuable, for the revision.

[1] [8].

[2] [3]. Cf. also [1], [7].

[3] [2], [5].

[4] Theorem 2 in § 7 below.

[5] [9].

[6] Together with every ring which we shall consider, except perhaps $\mathfrak{A}$ below. Further, we consider only those subrings of R which contain the unit element of R.

[7] [8], § 1; or [9], § 0. A right-module $\mathfrak{m}$ of R is called *regular*, when a direct sum of a certain number, say m, of its copies is isomorphic to the direct sum of a certain number, say n $(\neq 0)$, of copies of R itself, that is, $\mathfrak{m}^m \simeq R^n$. The number $k = n/m$, uniquely determined by $\mathfrak{m}$, is called its rank. We have: α) a right-regular module $\mathfrak{m}$ of R is also regular with respect to the (R-)endomorphism ring R^* with rank k^{-1}; β) the R^*-endomorphism ring of $\mathfrak{m}$ coincides with R; γ) if $\mathfrak{m}$ of rank k,

$V_{\mathfrak{A}}(S_l) = \mathfrak{B}$. If, conversely, S is a subring of R and R is right-regular with respect to S_l (that is, R is left-regular with respect to S), then $\mathfrak{B} = V_{\mathfrak{A}}(S_l)$ is a subring of $\mathfrak{A}$, containing R_r, with respect to which R is regular. *So the* $\mathfrak{B}$ *and* S *of these types correspond to each other* 1-1 *dually.* The R_r-rank of $\mathfrak{B}$ is equal to the S_l-rank of R.

Another notion which the writer wants to take from his former paper is that of a *Galois class-group.*[8] Consider a finite group of automorphism-classes of R, class referring to the inner automorphism group. Let $\rho_1, \rho_2, \cdots, \rho_g$ be its representative system. To each $\rho = \rho_i$ we may define an R-double-module (ρ, R) which is R itself as R-right-module and for which the left-operation of an element a in R is explained by the left-multiplication of a^ρ, that is, by $(a^\rho)_l$; it is determined uniquely up to isomorphism by the class of ρ. We say that our finite group of automorphism-classes, represented by $\rho_1, \rho_2, \cdots, \rho_g$, forms a Galois class-group if no (ρ_i, R) and (ρ_j, R) with $i \neq j$ possess mutually R-two-sided isomorphic composition residue-moduli.[9] In this case, each submodule of the direct sum $(\rho_1, R) + (\rho_2, R) + \cdots + (\rho_g, R)$ is a direct sum of submoduli of (ρ_i, R), $i = 1, 2, \cdots, g$, according to the well known theorem of the decomposition of a group into *idealistic* components.[10]

1. Now let $\mathfrak{B}_0$ be a subring of $\mathfrak{A}$ containing R_r and satisfying the minimum condition. Let further $G = \{1, \sigma, \cdots, \tau\}$ be a finite group of order g, and $\rho(1), \rho(\sigma), \cdots, \rho(\tau)$ automorphisms of R such that $\rho(\sigma)\rho(\tau)\rho(\sigma\tau)^{-1} \varepsilon \mathfrak{B}_0$ for every pair σ, τ.[11] Each $\rho(\tau)$, as well as any automorphism ρ of R, induces an automorphism in $\mathfrak{A}$ in natural manner: $x^\rho \alpha^\rho = (x\alpha)^\rho$ $(x \varepsilon R)$, or $\alpha^\rho = \rho^{-1}\alpha\rho$. We have $(a_l)^\rho = (a^\rho)_l$, $(a_r)^\rho = (a^\rho)_r$ (which depend essentially on the ring-automorphism property of ρ). Assume that in this sense $\rho(1), \rho(\sigma), \cdots, \rho(\tau)$ induce automorphisms in $\mathfrak{B}_0$, that is, $\rho(\sigma)^{-1}\mathfrak{B}_0\rho(\sigma) = \mathfrak{B}_0$ for every $\sigma \varepsilon G$, and moreover that $\{\rho(1), \rho(\sigma), \cdots, \rho(\tau)\}$ forms, again in the

---

is also regular and of rank l with respect to a subring T of R, then any other regular module of R is regular with respect to T too and its T-rank is equal to lk^{-1} times its R-rank.

[8] [8], § 3.

[9] If R is a division ring, or more generally a simple ring, this condition is implied automatically from the fact that we are dealing with a group of automorphism-*classes.*

[10] A direct decomposition of a group is called *idealistic,* if no two distinct components have isomorphic composition factor-groups and if the decomposition is the finest as such. We propose to use "idealistic" (rather than "ideal") not to be mistaken for "ideal" in a ring.

[11] $\rho(\sigma)\rho(\tau)\rho(\sigma\tau)^{-1}$ are regular elements in $\mathfrak{B}_0$, since they are non-zero-divisors, and $\mathfrak{B}_0$ is a ring with unit element and minimum condition.

same sense, a representative system of a Galois class-group of the ring $\mathfrak{B}_0$ (rather than R).

Let $\mathfrak{B}$ be the ring of endomorphisms of R generated by $\rho(1), \rho(\sigma), \cdots, \rho(\tau)$ and $\mathfrak{B}_0$, that is,

$$\mathfrak{B} = \rho(1)\mathfrak{B}_0 + \rho(\sigma)\mathfrak{B}_0 + \cdots + \rho(\tau)\mathfrak{B}_0.$$

Here $\rho(1)\mathfrak{B}_0, \rho(\sigma)\mathfrak{B}_0, \cdots, \rho(\tau)\mathfrak{B}_0$ are $\mathfrak{B}_0$-two-sided isomorphic to the moduli $(\rho(1), \mathfrak{B}_0), (\rho(\sigma), \mathfrak{B}_0), \cdots, (\rho(\tau), \mathfrak{B}_0)$ in our above notation, $\rho(1), \rho(\sigma), \cdots, \rho(\tau)$ being viewed as automorphisms of $\mathfrak{B}_0$. Hence our assumed Galois class-group property implies that the sum is direct, and $\mathfrak{B}$ is a so-called crossed product of $\{\rho(1), \rho(\sigma), \cdots, \rho(\tau)\}$ and $\mathfrak{B}_0$. Further, [8], Lemma 5 [11] can be applied to $\mathfrak{B}$. Namely, if two right-moduli of $\mathfrak{B}$ are $\mathfrak{B}_0$-regular and mutually $\mathfrak{B}_0$-isomorphic, then they are $\mathfrak{B}$-isomorphic.

Now suppose that R is regular, of rank $b = n/m$, with respect to $\mathfrak{B}_0$. Then the $\mathfrak{B}$-module R^{mg} is ($\mathfrak{B}_0$-, whence) $\mathfrak{B}$-isomorphic to $\mathfrak{B}^n$, by virtue of our fundamental lemma just cited, and R is $\mathfrak{B}$-regular with rank bg^{-1}. Thus § 0 can be applied to $\mathfrak{B}$; $V_{\mathfrak{A}}(\mathfrak{B}) = S_l$ $(S \subseteq R)$, $V_{\mathfrak{A}}(S_l) = \mathfrak{B}$ and R is S_l-regular with rank $b^{-1}g$. Hence we have

THEOREM 1. *Let* $\mathfrak{B}_0$ *be a subring (satisfying minimum condition) of the absolute endomorphism ring* $\mathfrak{A}$ *of* R *containing* R_r *such that* R *is* $\mathfrak{B}_0$-*regular of rank* b. *Let* $G = \{1, \sigma, \cdots, \tau\}$ *be a finite group and let a system* $\{\rho(1), \rho(\tau), \cdots, \rho(\tau)\}$ *of (ring-)automorphisms of* R *be such that* $\rho(\sigma)^{-1}\mathfrak{B}_0\rho(\sigma) = \mathfrak{B}_0$, $\rho(\sigma)\rho(\tau)\rho(\sigma\tau)^{-1} \varepsilon \mathfrak{B}_0$ *for every* $\sigma, \tau \varepsilon G$ *and form a representative system of a Galois class-group of* $\mathfrak{B}_0$. *Then* R *is regular of rank* bg^{-1} *with respect to* $\mathfrak{B} = \Sigma\rho(\sigma)\mathfrak{B}_0$, *and the commuter system* $V_{\mathfrak{A}}(\mathfrak{B})$ *(i. e., the* $\mathfrak{B}$-*endomorphism ring of* R*) is* S_l *with a subring* S *of* R *such that* R *is* S-*left-regular over* S *with rank* $b^{-1}g$, g *being the order of* G. *Further* $V_{\mathfrak{A}}(S_l) = \mathfrak{B}$.

2. (Our above Galois class-group condition, on the system $\{\rho(1), \rho(\sigma), \cdots, \rho(\tau)\}$ refers to $\mathfrak{B}_0$ rather than to R. Now we want to replace it by an assumption which refers to the latter, i. e. to R itself.) Let $\mathfrak{B}_0$ be a

---

[12] Namely: let X be a ring with unit element and minimum condition and let Y be its radical. Let a group $\mathfrak{G}$ of automorphism-classes of X induce a Galois class-group in the residue-ring X/Y (which is weaker that it forms one of X itself). Consider a crossed product $(\mathfrak{G}, X)$ (with factor set) and two $(\mathfrak{G}, X)$-right-moduli $\mathfrak{s}$, $\mathfrak{t}$, which are direct sums of X-submoduli X-isomorphic to directly indecomposable right-ideal components of X. If then $\mathfrak{s}/\mathfrak{s}Y$ and $\mathfrak{t}/\mathfrak{t}Y$ *are* X-*isomorphic*, $\mathfrak{s}$ *and* $\mathfrak{t}$ *are* $(\mathfrak{G}, X)$-*isomorphic.* Note that the residue-moduli $\mathfrak{s}/\mathfrak{s}Y$, $\mathfrak{t}/\mathfrak{t}Y$ in the assumption are replaced by the moduli $\mathfrak{s}$, $\mathfrak{t}$ themselves in the conclusion, and which is more important, the operator domain X by $(\mathfrak{G}, X)$. For a special case cf. [7], Lemma 1.

subring of $\mathfrak{A}$ which contains R_r and satisfies the double chain as R_r-two-sided module. Assume that every composition residue-module of $\mathfrak{B}_0$, as R_r-two-sided module, is isomorphic to one of R_r. Let a finite system $\{\rho(1), \rho(\sigma), \cdots, \rho(\tau)\}$ of automorphisms of R, given in connection with a group $G = \{1, \sigma, \cdots, \tau\}$, be such that $\rho(\sigma)^{-1}\mathfrak{B}_0\rho(\sigma) = \mathfrak{B}_0$, $\rho(\sigma)\rho(\tau)\rho(\sigma\tau)^{-1} \varepsilon \mathfrak{B}_0$, again, and moreover that the R_r-two-sided moduli $\rho(\sigma)R_r$, $\rho(\tau)R_r$ with distinct σ, τ have no isomorphic composition residue-moduli. Then the system $\{\rho(\tau)\}$ certainly forms a representative system of a Galois class-group of $\mathfrak{B}_0$, $\rho(\sigma)$ being viewed as automorphisms of $\mathfrak{B}_0$.

3. If in particular $\mathfrak{B}_0 \subseteq R_rR_l$ and $\mathfrak{B}_0$ is generated by a finite number of elements in R_l over R_r, then the condition we imposed on $\mathfrak{B}_0$ in the preceding section is fulfilled. We have $\mathfrak{B}_0 = R_rK_l$ with a subring K of R; indeed, with $K_l = \mathfrak{B}_0 \cap R_l$.[13] Moreover $K_l = (\Sigma\rho(\sigma)\mathfrak{B}_0) \cap R_l$, too,[14] where of course $\{\rho(\sigma)\}$ is as in § 2. The subring S of R given by $S_l = V_{\mathfrak{A}}(\mathfrak{B}) = V_{\mathfrak{A}}(\Sigma\rho(\sigma)\mathfrak{B}_0)$ is characterized as the totality of elements of R which are invariant under $\{(\rho), \rho(\sigma), \cdots, \rho(\tau)\}$ and which are commutative with K elementwise. For,

$$\begin{aligned} V_{\mathfrak{A}}(\mathfrak{B}) &= V_{\mathfrak{A}}(R_r) \cap V_{\mathfrak{A}}(K_l) \cap V_{\mathfrak{A}}(\{\rho(1), \rho(\sigma), \cdots, \rho(\tau)\} \\ &= R_l \cap V_{\mathfrak{A}}(K_l) \cap V_{\mathfrak{A}}(\{\rho(1), \rho(\sigma), \cdots, \rho(\tau)\}) \\ &= V_{R_l}(K_l) \cap V_{\mathfrak{A}}(\{\rho(1), \rho(\sigma), \cdots, \rho(\tau)\}). \end{aligned}$$

If, further R is regular with respect to $\mathfrak{B}_0 = R_rK_l$, then ($\mathfrak{B} = V_{\mathfrak{A}}(S_l)$ and so $K_l = \mathfrak{B}_0 \cap R_l = \mathfrak{B} \cap R_l = V_{\mathfrak{A}}(S_l) \cap R_l = V_{R_l}(S_l)$ whence) $K = V_R(S)$. We state these facts as

LEMMA 1. *Let a subring $\mathfrak{B}_0$ of $\mathfrak{A}$ contain R_r, satisfy the R_r-two-sided double chain condition, and moreover be such that every composition residue-module of $\mathfrak{B}_0$, as R_r-two-sided module, is isomorphic to one of R_r. Let a system $\{\rho(\tau)\}$ of automorphisms of R, with a finite group $G = \{\sigma\}$, be such that $\rho(\sigma)^{-1}\mathfrak{B}_0(\sigma) = \mathfrak{B}_0$, $\rho(\sigma)\rho(\tau)\rho(\sigma\tau)^{-1} \varepsilon \mathfrak{B}_0$, for all σ, $\tau \varepsilon G$, and that $(\rho(\sigma), R)$ and $(\rho(\tau), R)$ with distinct σ, τ, in the notation of* § 0, *have no isomorphic composition residue-moduli. Assume further that R is $\mathfrak{B}_0$-regular. Then Theorem* 1 *may be applied to $\mathfrak{B}_0$ and $\{\rho(\sigma)\}$. Our assumption on the R_r-behavior of $\mathfrak{B}_0$ is satisfied in particular if $\mathfrak{B}_0$ is contained in R_rR_l and is generated by a finite number of elements in R_l over R_r. If here we put $K_l = \mathfrak{B}_0 \cap R_l$, then $\mathfrak{B}_0 = R_rK_l$, and $\rho(\sigma)^{-1}\mathfrak{B}_0\rho(\sigma) = \mathfrak{B}_0$ is equivalent to*

---

[13] K contains the center of R.

[14] Cf. again the theorem of idealistic components.

$K^{\rho(\sigma)} = K$. *Moreover* $K = V_R(S)$, *where* S *is as in Theorem* 1 *and is in fact the totality of element of* $V_R(K)$ *invariant under* $\rho(1), \rho(\sigma), \cdots, \rho(\tau)$.

If we suppose that a finite set $a_1, a_2, \cdots, a_k$ of elements of K can be so chosen that the left-multiplications $a_{1l}, a_{2l}, \cdots, a_{kl}$, on R, form an independent basis of $\mathfrak{B}_0 = R_r K_l$ over R_r, then $a_1, a_2, \cdots, a_k$ form a basis of K over the center Z of R. For, an element $x_{1r}a_{1l} + x_{2r}a_{2l} + \cdots + x_{kr}a_{kl}$ $(x_i \varepsilon R)$ belongs to R_l if and only if every x_i is in Z. The converse is also valid as we readily see by a similar argument. And, if such is the case, $\mathfrak{B}_0$ is certainly right- (as well as left-) regular over R_r, with rank k. Hence, if R is $\mathfrak{B}_0$-right-regular, its rank b is k^{-1}, and R^{kg} is $\mathfrak{B}$-isomorphic to $\mathfrak{B}$, provided of course that our additional conditions are fulfilled. R possesses an independent left-basis over S with kg terms.

4. Let $\mathfrak{B}_0$, $\rho(\sigma)$ be as in § **1**, Theorem 1. Consider an endomorphism ϕ of R, as module, which commutes elementwise with $S_l : \phi \, \varepsilon \, V_{\mathfrak{A}}(S_l) = \mathfrak{B} = \Sigma\rho(\sigma)\mathfrak{B}_0$. If $\phi\mathfrak{B}_0 = \mathfrak{B}_0\phi$ and this $\mathfrak{B}_0$-two-sided module is directly indecomposable, then it is contained in an idealistic component of (the $\mathfrak{B}_0$-two-sided module) $\mathfrak{B}$, whence in one of $\rho(\sigma)\mathfrak{B}_0$. Here the second assumption is satisfied, either when ϕ is an automorphism of R as module and $\mathfrak{B}_0$ is (two-sided) directly indecomposable, or when $\mathfrak{B}_0$ is primary, provided that the first assumption $\phi\mathfrak{B}_0 = \mathfrak{B}_0\phi$ is fulfilled.

Suppose next that $\mathfrak{B}_0 = R_r K_l$ $(K_l = R_l \cap \mathfrak{B}_0)$ as in § **3** and that ϕ is an automorphism of the ring R (as a ring) which commutes with S_l elementwise, or, which is the same, leaves S elementwise invariant. Then $K^\phi = K$, since $K = V_R(S)$, and $\phi^{-1}K_l\phi = K_l$. Thus $\phi^{-1}\mathfrak{B}_0\phi = \mathfrak{B}_0$ and the former of our conditions, $\phi\mathfrak{B}_0 = \mathfrak{B}_0\phi$, is satisfied automatically. Thus

LEMMA 2. *Let* $\mathfrak{B}_0$, $\rho(\sigma)$, S *be as in Theorem* 1. *Let* $\phi \, \varepsilon \, \mathfrak{B} = V_{\mathfrak{A}}(S_l)$. *If* $\phi\mathfrak{B}_0 = \mathfrak{B}_0\phi$ *and* $\phi\mathfrak{B}_0$ *is* $\mathfrak{B}_0$*-two-sided directly indecomposable, then* ϕ *is contained in one of the* $\rho(\sigma)\mathfrak{B}_0$. *Here the second part of the last assumption, i. e. the indecomposability of* $\phi\mathfrak{B}_0$, *is satisfied either when* $\mathfrak{B}_0$ *is two-sided directly indecomposable and* ϕ *is an automorphism of* R *as module, or when* $\mathfrak{B}_0$ *is primary, provided that the first part is fulfilled. Further, if* $\mathfrak{B}_0 = R_r K_l$; *as in the second half of Lemma* 1, *then automatically* $\phi\mathfrak{B}_0 = \mathfrak{B}_0\phi$ *for every automorphism* ϕ *of* R *as ring leaving* S *elementwise invariant.*

Now, let Φ and Φ_0 be the totalities of automorphisms of R (as a ring) which are contained in $\mathfrak{B}$ and $\mathfrak{B}_0$, respectively. They are groups, since $\mathfrak{B}$ and $\mathfrak{B}_0$ are rings with unit element and satisfying minimum condition. If $\mathfrak{B}_0 = R_r K_l$ $(K_l = R_l \cap \mathfrak{B}_0)$ then Φ_0 is invariant in Φ, because $\phi K_l = K_l\phi$,

$\phi\mathfrak{B}_0 = \mathfrak{B}_0\phi$ for every $\phi \varepsilon \Phi$, as was just observed. If further $\mathfrak{B}_0 = R_rK_l$ is two-sided directly indecomposable, then every $\phi \varepsilon \Phi$ is contained in a certain $\rho(\sigma)\mathfrak{B}_0$, whence in $\rho(\sigma)\Phi_0$. So

LEMMA 3. *If again* $\mathfrak{B}_0 = R_rK_l$ *(and* R *is* $\mathfrak{B}_0$*-regular) then* Φ_0 *is an invariant subgroup of the group* Φ. *If further* $\mathfrak{B}_0 = R_rK_l$ *is two-sided directly indecomposable,* $\{\rho(\sigma)\}$ *forms a representative system of cosets in* Φ *with respect to the invariant subgroup* Φ_0, *and the factor-group* Φ/Φ_0 *is isomorphic to* G.

Let it be also observed that if K has the property that every element of K may be expressed as a sum of regular elements, then $R_rK_l = R_r\Phi_0$ [15] and indeed $R_rK_l = R_r\Phi_1$ where Φ_1 is the totality of inner automorphisms $a_ra_l^{-1}$ of R generated by regular elements a in K. This last property is enjoyed by a very wide class of rings [16]; we note here for later use, that primary rings certainly have the proposed property.

5. Now we wish to study a correspondence between certain subrings of $\mathfrak{B}$ and those of R. Let $\mathfrak{B}_0 = R_rK_l$, $\{\rho(\sigma)\}$ and S be as in § **2, 3.** Let $H = \{1, \alpha, \cdots, \beta\}$ be a subgroup of G, and let there be given for each $\alpha \varepsilon H$ an automorphism $\rho_1(\alpha)$ of R contained in $\rho(\alpha)\mathfrak{B}_0$ (whence in $\rho(\alpha)\Phi_0$). Let further $\mathfrak{B}_1$ be a subring of $\mathfrak{B}_0$ such that $\mathfrak{B}_1 \supseteq R_r$, $\rho_1(\alpha)^{-1}\mathfrak{B}_1\rho_1(\alpha) = \mathfrak{B}_1$, $\rho_1(\alpha)\rho_1(\beta)\rho_1(\alpha\beta)^{-1} \varepsilon \mathfrak{B}_1$ for all $\alpha, \beta \varepsilon H$. Obviously the ring $\mathfrak{B}^{\mathfrak{H}} = \Sigma\rho_1(\alpha)\mathfrak{B}_1$ is contained in $\mathfrak{B} = \Sigma\rho(\sigma)\mathfrak{B}_0$, and we have $R \supseteq S^{\mathfrak{H}} \supseteq S$, where $(S^{\mathfrak{H}})_l = V_{\mathfrak{A}}(\mathfrak{B}^{\mathfrak{H}})$. If here R is regular with respect to $\mathfrak{B}_1$,[17] then it is regular with respect to $\mathfrak{B}^{\mathfrak{H}}$ and $S^{\mathfrak{H}}{}_l$, and we have $\mathfrak{B}^{\mathfrak{H}} = V_{\mathfrak{A}}(S^{\mathfrak{H}}{}_l)$. Now the following lemma may be applied to $\mathfrak{C}_0 = \mathfrak{B}_1$ provided its supplementary conditions are fulfilled: [18]

LEMMA 4. *Assume that* $\mathfrak{B}_0 = R_rK_l$ *has an independent basis over* R_r *consisting of elements in* K_l. *Let* $\mathfrak{C}_0$ *be an* R_r*-two-sided submodule of* $\mathfrak{B}_0$ *such that* $\mathfrak{C}_0N_r = \mathfrak{C}_0 \cap N_rK_l$, *where* N *denotes the radical of* R, *and that* $\mathfrak{C}_0/\mathfrak{C}_0 \cap N_rK_l$ *is* R_r/N_r*-right-regular.*[19] *Then* $\mathfrak{C}_0$ *is a direct summand of* $\mathfrak{B}$ *as* R_r*-two-sided module and has also an independent basis over* R_r *consisting*

---

[15] $R_r\Phi_0$ denotes (not the mere product but) the product-module.

[16] See [10], Hilfssatz 5. The assumption there may be weakened so as to include the case of rings which are not algebras.

[17] This is the case if $\mathfrak{B}_0$ is right-regular with respect to $\mathfrak{B}_1$.

[18] We shall indeed apply the lemma to $\mathfrak{U}_\sigma$, $\mathfrak{U}^*_\sigma$ in § 6, 7, which satisfy our additional requirements.

[19] Under the first assumption, $\mathfrak{C}_0 \cap N_rK_l = \mathfrak{C}_0N_r$, this is of course equivalent to the statement that $\mathfrak{C}_0/\mathfrak{C}_0N_r$ is R_r/N_r-right-regular, and this last is certainly the case when $\mathfrak{C}_0$ is R_r-right-regular.

of elements of K_l. (*If* R *is primary, the condition* $\mathfrak{E}_0 N_r = \mathfrak{E}_0 \cap N_r K_l$ *is satisfied by every* R_r*-right-regular* $\mathfrak{E}_0$.)

This can be seen as in [9], § **3**. For the sake of completeness we repeat the argument briefly. First $\mathfrak{E}_0 \bmod N_r K_l$ (i. e. mod $\mathfrak{E}_0 \cap N_r K_l$) is an R_r/N_r-two-sided submodule of $R_r K_l / N_r K_l$. This last residue-module possesses an independent basis over R_r/N_r commutative with R_r/N_r elementwise, and so is a direct sum of R_r/N_r-two-sided moduli isomorphic with simple components of R_r/N_r. Hence $\mathfrak{E}_0/\mathfrak{E}_0 \cap N_r K_l$ possesses a similar structure. But the numbers of its components isomorphic with simple components of R_r/N_r are all equal, by virtue of its R_r/N_r-regularity. Hence $\mathfrak{E}_0/\mathfrak{E}_0 \cap N_r K_l$ is also decomposed into R_r/N_r-two-sided submoduli isomorphic to R_r/N_r and so possesses an independent basis over R_r/N_r commutative elementwise with R_r/N_r. Let the basis classes be represented by $\gamma_1, \gamma_2, \cdots, \gamma_t$ $(\varepsilon\, \mathfrak{E}_0)$; $\mathfrak{E}_0 \equiv \gamma_1 R_r + \gamma_2 R_r + \cdots + \gamma_t R_r \bmod N_r K_l$. Since, however, $\mathfrak{E}_0 \cap N_r K_l =$ (whence $\subseteq$) the intersection $\mathfrak{E}_0 N_r$ of all maximal R_r-right-submoduli by assumption, we have $\mathfrak{E}_0 = \gamma_1 R_r + \cdots + \gamma_t R_r$ too. (The sum is direct as we shall see in a moment.) Further, there exists a system $a_1, a_2, \cdots, a_s$ of elements in K such that $\gamma_1, \cdots, \gamma_t, a_{1l}, \cdots, a_{sl} \bmod N_r K_l$ together form an independent R_r/N_r-basis of $R_r K_l / N_r K_l$. We see, by an argument similar to the above, that

$$\begin{aligned} R_r K_l &= \gamma_1 R_r + \cdots + \gamma_t R_r + a_{1l} R_r + \cdots + a_{sl} R_r \\ &= \mathfrak{E}_0 + a_{1l} R_r + \cdots + a_{sl} R_r. \end{aligned}$$

Here the sums are direct as can be seen by consideration of R_r-lengths. Thus $\mathfrak{E}_0$ is a direct summand in $R_r K_l$ as R_r-two-sided module, and so is a direct sum of submoduli (R_r-two-sided) isomorphic to directly indecomposable two-sided ideal components of R_r. On observing that $\mathfrak{E}_0/\mathfrak{E}_0 N_r$ is regular, we see then that $\mathfrak{E}_0$ is a direct sum of submoduli R_r-two-sided isomorphic to R_r, which proves the lemma.

6. We next start conversely with a ring T between R and S; $R \supseteq T \supseteq S$. We have

$$R_r = V_{\mathfrak{A}}(R_l) \subseteq V_{\mathfrak{A}}(T_l) \subseteq V_{\mathfrak{A}}(S_l) = \mathfrak{B} = \Sigma \rho(\sigma) \mathfrak{B}_0.$$

Hence $V_{\mathfrak{A}}(T_l) = \Sigma \rho(\sigma) \mathfrak{U}_\sigma$ for certain R_r-two-sided submoduli $\mathfrak{U}_\sigma$ of $\mathfrak{B}_0$; observe the Galois class-group property of the system $\{\rho(\sigma)\}$.[20] In particular $\mathfrak{U}_1 \supseteq R_r$.

---

[20] The same is the case for every R_r-two-sided submodule of $\mathfrak{B}$ and in particular for every subring of $\mathfrak{B}$ containing R_r.

Assume now that R is (T_l-right-, that is) T-left-regular. Then T is S-left-regular. Naturally R_r is T_r-left-regular and R_l is T_l-right-regular. Further, $V_{\mathfrak{A}}(T_l)$ is R_r $(= V_{\mathfrak{A}}(R_l))$-right-regular. Hence each $\mathfrak{U}_\sigma$ is a direct sum of submoduli R_r-right-isomorphic to directly indecomposable right-ideal direct components of R. Concerning these R_r-right-components of $\mathfrak{U}_\sigma$ we have

LEMMA 5. *Under our assumption that R is T-left-regular, no direct component of* $\mathfrak{U}_\sigma$ *as* R_r-*right-module is contained in* $\mathfrak{B}_0 N_r$, *whence* $\mathfrak{U}_\sigma \cap \mathfrak{B}_0 N_r = \mathfrak{U}_\sigma N_r$, *or, which is the same,* $\mathfrak{U}_\sigma \cap \mathfrak{B}_0 N_r$ *coincides with the intersection of all the maximal* R-*right-submoduli of* $\mathfrak{U}_\sigma$, N *being the radical of* R.

To prove this, let[21] the $\mathfrak{B}_0$-right-rank of R be $b = n/m$, with integral n, m. The R_r-right-rank of $\mathfrak{B}_0$ is then $b^{-1} = m/n$. Denote the T-left-rank of R by $h = w/v$, with integers w, v. T is S-left-regular with rank $b^{-1}gh^{-1} = mgv/nw$. Hence $V_{\mathfrak{A}}(S_l) = \mathfrak{B} = \Sigma\rho(\sigma)\mathfrak{B}_0$ is $V_{\mathfrak{A}}(T_l)$-right-regular with rank mgv/nw. We consider the direct sum of nw copies of $\mathfrak{B}$:

$$\mathfrak{B}^{nw} = x^{(1)}\mathfrak{B} + x^{(2)}\mathfrak{B} + \cdots + x^{(nw)}\mathfrak{B}.$$

Let R be a direct sum of r directly indecomposable right-ideals (and left-ideals too). Our module $\mathfrak{B}^{nw}$ is a direct sum of $wmgr$ directly indecomposable R_r-right-moduli. We consider the module modulo the submodule $\mathfrak{B}^{nw}N_r = (\mathfrak{B}N_r)^{nw}$. The residue-module is decomposed into $wmgr$ irreducible R_r-right-moduli.

On the other hand, our module $\mathfrak{B}^{nw}$ is ($V_{\mathfrak{A}}(T_l)$-, whence) R_r-isomorphic to $V_{\mathfrak{A}}(T_l)^{mgv}$. So, if $\mathfrak{U}_\sigma$ is decomposed into t_σ directly indecomposable R_r-right-moduli, then we must have $mgv\Sigma t_\sigma = wmgr$, or, $v\Sigma t_\sigma = wr$. We assert that none of these t_σ, for each direct component of $\mathfrak{U}_\sigma$, is contained in $\mathfrak{B}_0 N_r$. Indeed, if such a component, say $\mathfrak{u}_\sigma$, were contained in $\mathfrak{B}_0 N_r$, then $\rho(\sigma)\mathfrak{u}_\sigma \subseteq \mathfrak{B}N_r$, and $\mathfrak{B}^{nw}$ $(\simeq V_{\mathfrak{A}}(T_l)^{mgv})$ mod $\mathfrak{B}^{nw}N_r$ would be a direct sum of a smaller number of irreducible moduli than $mgv\Sigma t_\sigma$ $(= wmgr)$,[22] which is a contradiction. Thus no direct (R_r-)component of $\mathfrak{U}_\sigma$ is contained in $\mathfrak{B}_0 N_r$, and the lemma is proved.

Our lemma holds without the assumption that $\mathfrak{B}_0$ has the form $R_r K_l$. If however it is assumed and if indeed $\mathfrak{B}_0$ has an independent R_r-basis consisting of elements of K_l, then the lemma says that Lemma 4, of the preceding section, can be applied to our $\mathfrak{U}_\sigma$ in place of $\mathfrak{C}_0$.

---

[21] Cf. [8], § 2.

[22] Observe that each component contributes at most one irreducible component to the residue-module.

7. Now let $\mathfrak{B}_0 = R_r K_l$ ($K_l = \mathfrak{B}_0 \cap R_l$) be two-sided directly indecomposable. Let Φ, Φ_0 be the totalities of (ring-)automorphisms of R contained in $\mathfrak{B}$, $\mathfrak{B}_0$, respectively, as in § 4. Consider a ring $\mathfrak{C}$ between $\mathfrak{B}$ and R_r; $\mathfrak{B} \supseteq \mathfrak{C} \supseteq R_r$. $\mathfrak{C}$ has the form $\mathfrak{C} = \Sigma \rho(\sigma)\mathfrak{U}_\sigma$. Put $\mathfrak{C} \cap \Phi = \Psi$ and $\mathfrak{C} \cap \Phi_0 = \Psi_0$. These are groups.[23] Choose a second representative system $\{\rho^*(\sigma)\}$ of $\Phi \bmod \Phi_0$ such that its subsystem, say $\{\rho^*(1), \rho^*(\alpha), \cdots, \rho^*(\beta)\}$, forms a representative system of $\Psi \bmod \Psi_0$. Clearly, $\{1, \alpha, \cdots, \beta\}$ is a subgroup of G, which we shall denote by H. Also $\mathfrak{B} = \Sigma \rho^*(\sigma)\mathfrak{B}_0$, and $\mathfrak{C} = \Sigma \rho^*(\sigma)\mathfrak{U}^*_\sigma$ with submoduli $\mathfrak{U}^*_0$ of $\mathfrak{B}_0$; here $\mathfrak{U}^*_\alpha \supseteq R_r$ for $\alpha \varepsilon H$.

If in particular $\mathfrak{C} = V_{\mathfrak{A}}(T_l)$ with $R \supseteq T \supseteq S$, and R is T-left-regular, then each $\mathfrak{U}^*_\sigma$ is a direct sum of R_r-right-moduli isomorphic with right-ideal direct components of R, and moreover $\mathfrak{U}^*_\sigma N_r = R_r K_l N_r \cap \mathfrak{U}^*_\sigma = N_r K_l \cap \mathfrak{U}^*_\sigma$, as was shown in the preceding section. If here R is primary, $\mathfrak{U}^*_\sigma$ is R_r-right-regular unless it vanishes. If further $R_r K_l$ has an independent basis over R_r, consisting of elements of K_l, then, by Lemma 4, each $\mathfrak{U}^*_\sigma$ has also an independent R_r-basis $\{a_{il}{}^{(\sigma)}\ (i = 1, 2, \cdots, u_\sigma)\}$ consisting of elements of K_l; $\mathfrak{U}^*_\sigma = \sum_{i=1}^{u_\sigma} R_r a_{il}{}^{(\sigma)}$ ($a_i{}^{(\sigma)} \varepsilon K$). Supposing moreover that all the $a_i{}^{(\sigma)}$ (with varying σ, i) are regular elements of K (which is certainly the case if $K \bmod K \cap N$ is a division ring),[24] we assert

Lemma 6. *Under all these assumptions,* $\mathfrak{C} = R_r\Psi = \sum_{\alpha \varepsilon H} \rho^*(\alpha) R_r \Psi_0$, *and in fact* $\mathfrak{U}^*_\sigma = R_r\Psi_0 = R_r L_l$ *whenever* $\mathfrak{U}^*_\sigma \neq 0$, *where* L *is the module spanned by* $a_1{}^{(1)}, a_2{}^{(1)}, \cdots$ *over the center* Z *of* R *and satisfies* $L_l = R_l \cap \mathfrak{C} = R_l \cap \mathfrak{U}^*_\sigma$, $L = V_R(T)$.

For, on taking σ with $\mathfrak{U}^*_\sigma \neq 0$, we have $\mathfrak{U}^*_\sigma \supseteq R_r a_{il}{}^{(\sigma)} \ni (a_i{}^{(\sigma)})_r{}^{-1} a_{il}{}^{(\sigma)}$ and $\rho^*(\sigma)(a_i{}^{(\sigma)})_r{}^{-1} a_{il}{}^{(\sigma)} \varepsilon \mathfrak{C} \cap \Phi = \Psi$. So $\sigma \varepsilon H$ and

$$(a_i{}^{(\sigma)})_r{}^{-1} a_{il}{}^{(\sigma)} = \rho^*(\sigma)^{-1} \rho^*(\sigma)\ (a_i{}^{(\sigma)})_r{}^{-1} a_{il}{}^{(\sigma)} \varepsilon \mathfrak{B}_0 \cap \Psi = \Psi.$$

Thus $R_r a_{il}{}^{(\sigma)} = R_r (a_i{}^{(\sigma)})_r{}^{-1} a_{il}{}^{(\sigma)} \subseteq R_r \Psi_0$ and $\mathfrak{U}^*_\sigma \subseteq R_r \Psi_0$. Therefore

$$\mathfrak{C} = \Sigma \rho^*(\sigma)\mathfrak{U}^*_\sigma \subseteq \sum_{\alpha \varepsilon H} \rho^*(\alpha) R_r \Psi_0.$$

Since the inclusion in the opposite direction is evident, we have

$$\mathfrak{C} = \sum_{\alpha \varepsilon H} \rho^*(\alpha) R_r \Psi_0 = R\Psi.$$

---

[23] $\mathfrak{C}$ satisfies, together with $\mathfrak{B}$, R_r-(whence $\mathfrak{C}$-) chain condition.

[24] Since $a_i(\sigma)$ are not contained in N anyway; if $a \varepsilon N$, then $N_r{}^t a_l = 0$ with $N^t \neq 0$.

So $\mathfrak{U}^*_\sigma = R_r\Psi_0 = \mathfrak{U}^*_1$ whenever $\mathfrak{U}^*_\sigma \neq 0$. Thus $\mathfrak{U}^*_\sigma = R_rL_l$ (for such σ). Further the cited properties of L are also evident.

Now we formulate our

THEOREM 2. *Let R be a primary ring (with unit element and satisfying minimum condition). Let Φ be a group automorphism of R which satisfies the following conditions:* i) *the subring K of R defined by $K_l = R_r\Phi \cap R_l$,*[25] *certainly containing the center Z of R, is regular over Z, or, what amounts to the same (since Z is completely primary), K has an independent (finite) basis over Z,* ii) *$K \bmod K \cap N$ is a division ring, where N denotes the radical of R,*[26] iii) *R is R_rK_l-regular,* iv) *every automorphism of R contained in R_rK_l is contained in Φ,* v) *the (necessarily) invariant subgroup $\Phi_0 = \Phi \cap R_rK_l$ has a finite index in Φ, and* vi) *there exists a representative system $\{\rho_i\}$ of $\Phi \bmod \Phi_0$ such that ρ_iR_r, ρ_jR_r with $i \neq j$ have no R_r-two-sided isomorphic composition residue-moduli.*[27, 28]

Let S be the invariant system of Φ in R, or, which is the same, the totality of elements of R commutative elementwise with K and invariant under $\{\rho_i\}$. Then Φ exhausts those automorphisms of R which leave S elementwise invariant. R has an S-left-basis of $(\Phi:\Phi_0)(K:Z)$ terms.

Let T be a ring between R and S: $R \supseteq T \supseteq S$, such that R is T-left-regular. Let Ψ be the totality of automorphisms of R leaving T elementwise fixed. Then T is the invariant system of Ψ in R. Ψ possesses a similar structure as Φ with a subring L of K in place of K, which satisfies the properties i)-vi) *with K, Φ, Φ_0 replaced by $L, \Psi, \Psi_0 = \Psi \cap R_rL_l$ $(= \Psi \cap R_rR_l)$. Further, not only is R T-left-regular but R has an independent T-left-basis of $(\Psi:\Psi_0)(L:Z)$ terms, and T has an S-left-basis of $(\Phi:\Phi_0)(K:Z)/(\Psi:\Psi_0)(L:Z)$ terms.*

If conversely Ψ is a subgroup of Φ such that L, defined by $L_l = R_r\Psi \cap R_l$, has an independent basis over Z,[29] *and every automorphism of R contained*

---

[25] See footnote 15.

[26] (K is then completely primary, whence directly indecomposable, and therefore) R_rK_l is (two-sided) directly indecomposable, since every direct summand of R_rK_l as R_r-two-sided module possesses an R_r-basis consisting of elements of K_l; of course i) being taken into account.

[27] That is, R-two-sided moduli (ρ_i, R), (ρ_j, R) with $i \neq j$, in the notation of § 0, have no isomorphic composition residue-moduli.

[28] Then $\Phi_0 = \Phi \cap R_rR_l$ too.

[29] Then we see that K has a right-, say, basis over L by an argument similar to one in our proof of Lemma 4, observing that $L \bmod N$ is a division ring. So (R_rK_l whence) R is R_rL_l-regular. We may also replace conversely our assumption by the L-right-, say, regularity of K.

in R_rL_l *is in* Ψ *(i. e. the properties* i), iv) *with* Ψ, L *in place of* Φ, K *are satisfied)*,[30] *then* R *has an independent left-basis over the invariant system* T *of* Ψ *(whence* R *is certainly* T*-left-regular), and* Ψ *exhausts the automorphisms of* R *leaving* T *elementwise invariant.*

Thus such between-rings T *(over which* R *is left-regular) and such subgroups* Ψ *(having the described structure) are in* 1-1 *dual correspondence.*

Proof is immediate by virtue of our preparations.

8. Our Theorem 2 includes (the main) Theorem 8 of [5] (and the corresponding statement in [2]) as a special case where R is a division ring, if the following fact is taken into consideration:

LEMMA 7.[31] *If* R *is a division ring and* K *a subring containing the center* Z *of* R*, then every automorphism of* R*, whose inverse is contained in* R_rK_l *is an inner automorphism generated by an element of* K*.*

To see this, first let an automorphism ϕ of R be contained in R_rR_l; $\phi-(t_{1r}s_{1l}+t_{2r}s_{2l}+\cdots)=0$, where $t, s\ (\neq 0)\,\varepsilon\, R$ and $t_{1r}, t_{2r}, \cdots$ are linearly independent over R_l (i. e. $t_1, t_2, \cdots$ are so over Z). For $x \,\varepsilon\, R$, $0=x_l\phi-x_l(t_{1r}s_{1l}+\cdots)-\phi x_l^{\phi}+(t_{1r}s_{1l}+\cdots)x_l^{\phi}=t_{1r}(s_{1l}x_l^{\phi}-x_ls_{1l})+\cdots$. Hence $s_{1l}x_l^{\phi}-x_ls_{1l}=0$, that is, $x^{\phi}s_1-s_1x=0$, $x^{\phi}=s_1xs_1^{-1}$, and ϕ is the inner automorphism $s_{1r}^{-1}s_{1l}$, generated by s_1^{-1}. Suppose now $\phi\,\varepsilon\, R_rK_l$. Then $s_{1l}\,\varepsilon\, s_{1r}R_rK_l=R_rK_l$, whence $\varepsilon\, R_rK_l \cap R_l=K_l$, i. e. $s_1\,\varepsilon\, K$, which proves our assertion.

We note that the somewhat complicated, though not irrational, assumptions in our Theorem, concerning K, become rather natural (and in fact some of them are satisfied automatically) also in case R is a simple ring,[32] and similarly in the case when K, instead of R, is simple.[33]

On the other hand, it is very desirable to replace the assumption of primarity[34] of R by mere direct indecomposability. Indeed we have done that in case $\Phi=1$. Namely: *Let* $\mathfrak{G}$ *be a Galois group*[35] *of a ring* R *(with*

---

[30] Besides ii), which is trivial, and iii), which was observed in the preceding footnote, properties v), vi) with Ψ, L in place of Φ, K are satisfied automatically. Cf. footnote 28.

[31] Cf. [5], Lemma 2, which in turn is contained in our argument of the Galois class-group and idealistic components.

[32] Cf. a remark at the end of [9], § 3.

[33] See [9], § 2, Example.

[34] As a matter of fact, in case of a primary ring R our Lemma 5 was superfluous, or evident from the regularity of $\mathfrak{U}_{\sigma}$, $\mathfrak{U}^*_{\sigma}$.

[35] That is, a finite group of automorphisms which forms by itself a representative system of a Galois class-group. See [8].

unit element and satisfying minimum condition), and let S be the invariant system of $\mathfrak{G}$ in R. (Then R is both left- and right-regular with rank $g=(\mathfrak{G}:1)$ over the invariant system S of $\mathfrak{G}$. If R is two-sided directly indecomposable, $\mathfrak{G}$ exhausts automorphisms of R leaving S elementwise invariant.) Under the same assumption of direct indecomposability of R, subgroups of $\mathfrak{G}$ are in 1-1 dual correspondence, in the usual manner of the Galois theory, with subrings of R, containing S, over which R is right- (or left-) regular. The only point which is needed to establish this and which is not covered by our argument in the present note is that of showing $\mathfrak{U}_\sigma = R_r$ for $\mathfrak{U}_\sigma \neq 0$ (when $\mathfrak{B}_0 = R_r$) (see § 6 above). We want to refer for it to the latter half of the proof given to Theorem 2 in [8].

The writer feels that the strict division into commuter and Galois class-group parts, so to speak, is not very natural in case R is not a division ring, or at least a simple ring. In all, he does not assume that the above is the final form of the theory; he hopes to come back to the problem elsewhere. A non-normal theory, which generalizes [4], will be given in a subsequent paper.

UNIVERSITY OF ILLINOIS.

REFERENCES.

[1] G. Azumaya, "Galois theory for uni-serial rings," *Journal of Mathematical Society of Japan*, vol. 1 (1949), pp. 130-137.

[2] H. Cartan, "Théorie de Galois pour les corps non-commutatifs," *Annales Scientifiques de l'École Normale Supérieure*, vol. 65 (1948), pp. 60-77.

[3] N. Jacobson, "The fundamental theorem of Galois theory for quasifields," *Annals of Mathematics*, vol. 41 (1940), pp. 1-7.

[4] ———, "An extension of Galois theory to non-normal and non-separable fields," *American Journal of Mathematics*, vol. 66 (1944), pp. 1-29.

[5] ———, "A note on division rings," *American Journal of Mathematics*, vol. 69 (1947), pp. 27-36.

[6] T. Nakayama and G. Azumaya, "On irreducible rings," *Annals of Mathematics*, vol. 48 (1947), pp. 949-965.

[7] T. Nakayama, "Semilinear normal basis for quasifields," *American Journal of Mathematics*, vol. 71 (1949), pp. 241-248.

[8] ———, "Galois theory for general rings with minimum condition," *Journal of Mathematical Society of Japan*, vol. 1 (1949), pp. 203-216.

[9] ———, "Commuter systems in a ring with radical," *Duke Mathematical Journal*, vol. 16 (1949), pp. 331-337.

[10] K. Shoda, "Ueber die Galoissche Theorie der halbeinfachen hyperkomplexen Systeme," *Mathematische Annalen*, vol. 107, pp. 252-258.

LINEARLY COMPACT SPACES AND DOUBLE VECTOR SPACES OVER SFIELDS.*

By J. Dieudonné.

Introduction. The notion of *linearly compact vector space* has recently been introduced by S. Lefschetz, in view of applications to Algebraic Topology ([7] p. 78). It is closely connected with the general theory of *weak duality* of vector spaces over topological sfields, which I have developed for algebraic purposes [2]; and I shall in fact show in Section 1 how the use of that theory greatly simplifies the proof of Lefschetz's main structure theorem. In the remainder of the paper, I will use linearly compact vector spaces to complete the recent generalization of Jacobson's theory of "composites" of self-representations of fields [5] developed by G. Hochschild [4]; it turns out that they are the proper tool by which Hochschild's finiteness restrictions may be entirely dropped.

1. Linearly compact spaces. Let K be an arbitrary sfield, on which the topology will always be the discrete topology. A right topological vector space E over K is said to be *linearly compact* when:

a) it is a Hausdorff space in which there is a fundamental system of neighborhoods of 0 consisting of linear subspaces of E;

b) any filter base on E consisting of closed linear varieties has a non-empty intersection.

Spaces satisfying a) are said to be *linearly topologized.* The following are easy consequences of the definition ([7], p. 78):

c) a product of linearly compact spaces is linearly compact;

d) if E is linearly compact, and H is a closed linear subspace of E, then H and E/H are linearly compact;

e) let E be linearly compact, F linearly topologized, and let u be a continuous linear mapping of E into F; then $u(E)$ is linearly compact, and u is a homomorphism[1] of E onto $u(E)$;

f) a discrete linearly compact space is finite dimensional.

Let E be a linearly compact space, E' its *dual* in the *topological* sense, namely the left vector space (over K) consisting of all *continuous* linear

---

* Received September 6, 1949.

[1] We use the term "homomorphism" in the sense of Bourbaki, meaning what is called "open homomorphism" by some authors.

mappings of E into K (continuous linear forms on E). Assumption a) implies that for every $x_0 \neq 0$ in E there exists a closed hyperplane H such that $x_0 \notin H$; this amounts to saying that there exists a continuous linear form $x' \in E'$ such that $x'(x_0) \neq 0$; E may therefore be considered as a subspace of the *dual* E'^* of E' in the *algebraic* sense, that is, the right vector space of *all* linear forms on E'. Now we prove the following properties:

g) *The topology on E is identical to the weak topology $\sigma(E, E')$* (topology of pointwise convergence in E'). Indeed, let V be an open (and closed) linear subspace of E; E/V is linearly compact by d), and discrete, therefore finite dimensional by f); hence V may be defined by a finite number of equations $x'_i(x) = 0$, with $x'_i \in E'$, and is therefore a neighborhood of 0 for $\sigma(E, E')$, which proves our contention.

h) *The linearly compact space E is complete.* Let $\mathfrak{F}$ be a Cauchy filter on E; for every open linear subspace V of E and every set $A \in \mathfrak{F}$ let us consider the set $A + V$; it is readily seen that these sets constitute a Cauchy filter. In fact, for every such open subspace V, there exists $A \in \mathfrak{F}$ such that for $x \in A$ and $y \in A$, $x - y \in V$, hence $A + V = x + V$, $A + V$ is a closed linear variety. Clearly the sets $A + V$ which are linear varieties form a filter base, hence have by assumption a common point x_0, which is the limit of the filter $\mathfrak{F}$.

The main structure theorem of Lefschetz is now an immediate consequence of g) and h) and of the well-known fact that E, with the topology $\sigma(E, E')$ may only be complete when $E = E'^*$ ([2], p. 66); E is thus the *algebraic dual* of a left vector space E' over K; if $(e_\lambda)_{\lambda \in L}$ is a basis for E', E is thus algebraically isomorphic to the *product vector space* K^L and the topology $\sigma(E, E')$ on E is identical to the product topology; this is Lefschetz's theorem ([7], p. 83, theorem (32.1)).

On the dual E' of E, Lefschetz considers the discrete topology (which is also the topology of uniform convergence in E); it will be more useful for us to consider the *weak topology* $\sigma(E', E)$ (topology of pointwise convergence in E); we stress that for that topology as well as for the discrete topology, *every* linear subspace in E' is *closed*.

2. Cyclic double modules and relations spaces. If A and B are two rings with identity elements, we consider (A, B)-modules, that is, sets M having simultaneously the structure of a left A-module and of a right B-module, with the conditions $1 \cdot x = x \cdot 1 = x$ and $(\alpha x)\beta = \alpha(x\beta)$ for $x \in M$, $\alpha \in A$, $\beta \in B$; such a module M is called *cyclic* if there exists an element $m \in M$ such that $M = A \cdot m \cdot B$ (set of all linear combinations $\sum_i \alpha_i m \beta_i$); m is then called a generator of M.

We shall be concerned with cyclic (K, A)-modules, where K is a *sfield* and A a ring with identity. Let M be such a module, m_0 a generator of M; M is a left vector space over K. We consider with Hochschild [4], the algebraic *dual* M' of that vector space; this is a right vector space over K, $x'\lambda$ being the linear form $x \to x'(x)\lambda$. On the other hand it may be considered as a left A-module, the product $u \cdot x'$ of $u \in A$ and $x' \in M'$ being the linear form $x \to x'(x \cdot u)$. The usual axioms are readily verified, as well as the condition $u \cdot (x'\lambda) = (u \cdot x')\lambda$ for $\lambda \in K$ and $u \in A$, and therefore M' is an (A, K)-module.

Next we associate to M and m_0 what Hochschild calls the *relations space* of M (relative to m_0). Let $\mathcal{E}$ be the additive group of *all* representations of the additive group A into the additive group K; to every element $x' \in M'$ we associate the element $\xi = \phi(x')$ of $\mathcal{E}$ defined by $\xi(u) = x'(m_0 \cdot u)$ for every $u \in A$; owing to the fact that m_0 is a generator of M, ϕ is a one-one mapping of M' onto a subgroup $R(M, m_0)$ of $\mathcal{E}$. Moreover, one may define on $\mathcal{E}$ a structure of (A, K)-module, by setting $\xi\lambda$ $(\lambda \in K)$ equal to the representation $u \to \xi(u)\lambda$, and $v \cdot \xi$ $(v \in A)$ to the representation $u \to \xi(u \cdot v)$; it is then clear that $R(M, m_0)$ is an (A, K)-submodule of $\mathcal{E}$, and ϕ an (A, K)-isomorphism of the (A, K)-module M' onto $R(M, m_0)$.

Now consider on M' the weak topology $\sigma(M', M)$, and on $\mathcal{E}$ the topology of pointwise convergence in A; for these topologies, ϕ is a *homeomorphism* of M' onto $R(M, m_0)$. In fact, every finite subset $(x_i)_{1 \leq i \leq n}$ of M is contained in a subspace (over K) generated by a finite number of elements of the form $m_0 \cdot u_j$ $(1 \leqq j \leqq p)$, and therefore the relations $x'(x_i) = 0$ (which define a neighborhood of 0 in M') are implied by $x'(m_0 \cdot u_j) = 0$, that is $(\phi(x'))(u_j) = 0$; this proves that ϕ maps a neighborhood of 0 in M' onto a neighborhood of 0 in $R(M, m_0)$; the same property for ϕ^{-1} is obvious.

The characterization of relations spaces $R(M, m_0)$ in $\mathcal{E}$, generalizing Hochschild's Theorem 5.3 (which concerns the finite case) now reads as follows:

THEOREM 1. *In order that an (A, K)-submodule T of $\mathcal{E}$ be the relations space of a cyclic (K, A)-module, it is necessary and sufficient that T, with the topology induced by that of $\mathcal{E}$, be a linearly compact right vector space over K.*

We observe first that an (A, K)-submodule T of $\mathcal{E}$ is always linearly topologized by the topology induced by that of $\mathcal{E}$, for a neighborhood V of 0 consists of all $\xi \in T$ such that $\xi(u_i) = 0$ for a finite number of $u_i \in A$, and it is clear that the relation $\xi \in V$ implies $\xi\lambda \in V$ for every $\lambda \in K$, which proves that V is a linear subspace of T.

The necessity of the condition on T is a consequence of the remarks preceding Theorem 1. To prove the sufficiency, we proceed as in the proof of Hochschild's Theorem 5. 3, and consider the *dual* $M = T'$ of the right vector space T (in the topological sense). M may be made into a (K, A)-module by setting λm $(\lambda \varepsilon K)$ equal to the mapping $\xi \rightarrow \lambda m(\xi)$ (this is the usual structure of left vector space of T'), and $m \cdot u$ $(u \varepsilon A)$ equal to the mapping $\xi \rightarrow m(u \cdot \xi)$. Let m_0 be the mapping $\xi \rightarrow \xi(1)$ of T into K, which is obviously a continuous linear form, and therefore belongs to M; we are going to prove that $M = K \cdot m_0 \cdot A$. It is clear that $K \cdot m_0 \cdot A$ is a linear subspace of the vector space $T' = M$; as all linear subspaces of T' are closed for the topology $\sigma(T', T)$, it will be enough to show that $K \cdot m_0 \cdot A$ is *dense* in T' for that topology. We use the following lemma of Hochschild ([4], Lemma 2. 1):

LEMMA 1. *Let S be an arbitrary set, V a linear subspace of finite dimension n in the right K-vector space K^S of all mappings of S into the sfield K; then there exist n elements s_i $(1 \leqq i \leqq n)$ of S and a basis $(\xi_i)_{1 \leq i \leq n}$ of V such that $\xi_i(s_j) = \delta_{ij}$.*[2]

Let then ξ_i $(1 \leqq i \leqq n)$ be n arbitrary elements of T, and V the linear subspace of T they generate; as $V \subset T \subset K^A$, there are p elements u_j $(1 \leqq j \leqq p)$ of A and a basis $(\eta_j)_{1 \leq j \leq p}$ of V such that $\eta_j(u_k) = \delta_{kj}$. If m is an arbitrary element of M, and $m' = \sum_{j=1}^{p} m(\eta_j) \cdot m_0 \cdot u_j$, we have $m'(\eta_k) = \sum_{j=1}^{p} m(\eta_j) m_0(u_j \cdot \eta_k)$, and $m_0(u_j \cdot \eta_k) = (u_j \cdot \eta_k)(1) = \eta_k(u_j) = \delta_{jk}$, hence $m'(\eta_k) = m(\eta_k)$ for $1 \leqq k \leqq p$; accordingly $m'(\xi_i) = m(\xi_i)$ for $1 \leqq i \leqq n$, which proves that $K \cdot m_0 \cdot A$ is dense in M, and therefore equal to M. Finally, T is the algebraic dual of M, when one sets $\xi(m) = m(\xi)$, and it follows from the definition of m_0 that $\xi(m_0 \cdot u) = \xi(u)$, hence the mapping ϕ is the identity, and we have proved that $T = R(M, m_0)$.

As a corollary to Theorem 1, we may prove Hochschild's theorems 6. 2 and 6. 3 *without any finiteness restriction.* We have first to show that, if M and N are cyclic (K, A)-modules with generators m_0 and n_0, then $R(M, m_0) + R(N, n_0)$ is the relations space of another cyclic (K, A)-module; it is enough to show, by Theorem 1, that if T and U are two linearly compact (A, K)-submodules of $\mathcal{E}$ then $T + U$ is also linearly compact. Now if we consider the cartesian product $T \times U$, which is linearly compact by property c) of Section 1, the mapping $(\xi, \eta) \rightarrow \xi + \eta$ of $T \times U$ onto the subspace $T + U$ of $\mathcal{E}$ is linear and continuous; therefore $T + U$ is linearly compact by property e) of Section 1.

---

[2] This may easily be deduced from the general theory of duality in vector spaces in the following way: for each $s \varepsilon S$, the mapping $\xi \rightarrow \xi(s)$ of V into K is a linear form $\bar{s}$ over V, and therefore $s \rightarrow \bar{s}$ is a mapping of S into the dual V' of V; moreover the image of S by that mapping generates V', for the relation $\xi(s) = 0$ for every $s \varepsilon S$ implies $\xi = 0$ in V. Therefore, there exist n elements s_i $(1 \leqq i \leqq n)$ such that the $\bar{s}_i$ constitute a basis of V'; the ξ_i are then the elements of the "dual basis" of V (see [1], p. 46).

Next, when M is decomposed into the direct sum of a finite number of cyclic (K, A)-modules M_i $(1 \leqq i \leqq s)$, the dual M' is a *direct topological sum* of the duals M'_i (isomorphic to the cartesian product of the M'_i); these are of course the only kinds of decomposition of the relations space $R(M, m_0)$ which must be considered. With this condition on the decompositions of $R(M, m_0)$, Hochschild's Theorem 6. 3 and its proof extend at once to the general case.

3. Tensor products of cyclic double modules. Let H, K be two sfields, A a ring with identity element. From a cyclic (H, K)-module M with generator m_0, and a cyclic (K, A)-module N with generator n_0, we construct a cyclic (H, A)-module L with generator $m_0 \otimes n_0$ as in Hochschild ([4], § 7). Theorem 7. 1 of Hochschild's paper may then be generalized as follows:

THEOREM 2. *The product* $R(M, m_0)R(N, n_0)$ (that is, the (A, H)-module generated by the elements $\xi \circ \eta$, where $\xi \,\varepsilon\, R(M, m_0)$ and $\eta \,\varepsilon\, R(N, n_0)$) *is dense in the relations space* $R(L, m_0 \otimes n_0)$.

We follow again closely Hochschild's proof. If M' and N' are the duals of M and N, we define for every $x' \,\varepsilon\, M'$ and $y' \,\varepsilon\, N'$, the linear form $x'y'$ over $M \otimes N$ (hence over L) by the relation $x'y'(m \otimes n) = x'(m \cdot y'(n))$; then it is readily seen that $\phi(x'y') = \phi(x') \circ \phi(y')$, and therefore $R(M, m_0)R(N, n_0)$ is contained in $R(L, m_0 \otimes n_0)$. The remainder of Hochschild's proof may be simplified by the following argument: as L is a linear subspace of the tensor product $M \otimes N$ (considered as a left vector space over H), every linear form over L may be extended to a linear form over $M \otimes N$; our proof will then be complete if we show that, given any finite number of elements z_k of $M \otimes N$, and an arbitrary linear form z' over $M \otimes N$, there is a sum of linear forms of the type $x'y'$ which takes the same value as z' at every z_k. We may suppose that the z_k are of the form $m_i \otimes n_j$, where the m_i are linearly independent in M (over H) and the n_j linearly independent in N (over K); for a finite subset of $M \otimes N$ is always contained in the linear subspace spanned by a finite number of such elements. Now let x'_i be a linear form over M such that $x'_i(m_h) = \delta_{ih}$, and let y'_j be a linear form over N such that $y'_j(n_k) = \delta_{jk}$; one has then $x'_i y'_j(m_h \otimes n_k) = \delta_{ih}\delta_{jk}$, and therefore the linear combination $\sum_{i,j} \lambda_{ij} x'_i y'_j$, where the λ_{ij} are arbitrary in H, takes the value λ_{ij} at every element $m_i \otimes n_j$, and may therefore have at those elements the same value as an arbitrary linear form over $M \otimes N$.

Let us now consider only the case when H is a subsfield of K and $A = K$; we then define as Hochschild ([4], Definition 7.2) the product of a cyclic (H, K)-module M by itself, and say that M is (algebraically) closed if it covers its product by itself. Hochschild's Theorem 7.2 is then valid *without any finiteness assumption* on M: for if $R(M, m_0)$ is a ring, and therefore contains $R(M, m_0)R(M, m_0)$, it contains also the closure of that product; hence M covers its product by itself; the converse is obvious.

To go further, we shall restrict ourselves to the case in which $H = K$. We need first the generalization of the Jacobson-Bourbaki theorem which was conjectured in [3], p. 163. The set $\mathcal{E}$ of all representations of the additive group K into itself is then a *ring* and we prove the following:

THEOREM 3. *Let T be a closed subring of $\mathcal{E}$, which at the same time is a (K, K)-submodule of $\mathcal{E}$. Let L be the subsfield of K consisting of all elements $u \varepsilon K$ such that $\xi(ux) = u\xi(x)$ for every $x \varepsilon K$ and every $\xi \varepsilon T$. Then T coincides with the ring of all L-linear mappings of K into itself.*[3]

As T is closed, all we have to prove is that T is *dense* in the ring A of all L-linear mappings of K into itself. This amounts to the following: let b_i $(1 \leqq i \leqq n)$ be any finite number of elements of K, linearly independent over L; we must show that the n elements $\xi(b_i)$ may assume *any* system of n values in K, for a suitable $\xi \varepsilon T$; in other words, we must prove that the linear mapping $\xi \rightarrow (\xi(b_i))_{1 \leq i \leq n}$ of the right vector space T (over K) into the right vector space K^n is a mapping *onto* K^n. Now this is exactly what Bourbaki's argument (in [1], p. 72) proves: for it shows that in the opposite case there would exist a system $(\lambda_i)_{1 \leq i \leq n}$ of n elements of L, not all 0, such that $\sum_{i=1}^{n} \lambda_i \xi(b_i) = 0$ for every $\xi \varepsilon T$, that is, $\xi(\sum_{i=1}^{n} \lambda_i b_i) = 0$ for all $\xi \varepsilon T$; but when $\xi(x) = 0$ for every $\xi \varepsilon T$, one has $x = 0$, for if $x \neq 0$ and $\xi \neq 0$, there exists $y \varepsilon K$ such that $\xi(y) \neq 0$, and that relation may be written $\eta(x) \neq 0$ with $\eta = (x^{-1}y) \cdot \xi \varepsilon T$. Therefore one would have $\sum_{i=1}^{n} \lambda_i b_i = 0$ against the hypothesis.

We recall that conversely, when L is any subsfield of K, and T is the (closed) ring of all L-linear mappings of K into itself then L is the set of all $u \varepsilon K$ such that $\xi(ux) = u\xi(x)$ for $x \varepsilon K$ and $\xi \varepsilon T$ ([6], p. 233, Theorem

---

[3] (*Added in the proofs*). Professors N. Jacobson and H. Cartan have kindly informed me that they have been aware of this result for some time. Cartan's proof is identical with mine. Jacobson's proof runs as follows: T is a right K-module, hence an *irreducible* ring of endomorphisms of K. It is therefore a *dense* ring of L-linear mappings of K into itself ([6], p. 232, th. 6); as it is closed, it consists of *all* such mappings.

7). There exists therefore a 1-1 correspondence between subsfields of K and closed subrings of $\mathcal{E}$ which are at the same time (K, K)-submodules.[4]

We may now give the characterization of all closed cyclic (K, K)-modules:

THEOREM 4. *If K is any sfield, every closed cyclic (K, K)-module is equivalent with a tensor product $K \otimes_L K$ of K with itself, with respect to a subsfield L of K. Any two such products $K \otimes_L K$, $K \otimes_H K$ are equivalent if and only if $H = L$.*

Let L be any subsfield of K, and (e_α) a basis of K, considered as a left vector space over L; then the elements $1 \otimes e_\alpha$ constitute a basis of $M = K \otimes_L K$ as a left vector space over K. For every element ξ of the relations space $R(M, 1 \otimes 1)$ and every $\lambda \varepsilon L$ one has $\xi(\lambda u) = x'(1 \otimes (\lambda u)) = x'(\lambda(1 \otimes u)) = \lambda x'(1 \otimes u) = \lambda \xi(u)$ for every $u \varepsilon K$; this proves at once that $R(M, 1 \otimes 1)$ is identical with the ring of *all* L-linear mappings of K into itself. In particular, the ring $\mathcal{E}$ is identical with the relations space of the (K, K)-module $K \otimes_P K$, where P is the *prime field* contained in K; therefore $\mathcal{E}$, considered as a right vector space over K, is *linearly compact*; hence every *closed* linear subspace of $\mathcal{E}$ is also linearly compact (Section **1**, property d)). Theorem 4 is then an immediate consequence of Theorems 1 and 3, and of Hochschild's Theorem 5. 2.

UNIVERSITY OF NANCY (FRANCE).

BIBLIOGRAPHY.

[1] N. Bourbaki, *Eléments de Mathématique, Algèbre*, chap. II (Paris, Hermann, 1947).

[2] J. Dieudonné, "Sur le socle d'un anneau et les anneaux simples infinis," *Bulletin de la Société mathématique de France*, vol. 70 (1942), pp. 46-75.

[3] J. Dieudonné, "La théorie de Galois des anneaux simples et semi-simples," *Commentarii Mathematici Helvetici*, vol. 21 (1948), pp. 154-184.

[4] G. Hochschild, "Double vector spaces over division rings," *American Journal of Mathematics*, vol. 71 (1949), pp. 443-460.

[5] N. Jacobson, "An extension of Galois theory to non-normal and non-separable fields," *American Journal of Mathematics*, vol. 66 (1944), pp. 1-29.

[6] N. Jacobson, "Structure theory of simple rings without finiteness assumptions," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 228-245.

[7] S. Lefschetz, *Algebraic Topology* (Am. Math. Soc. Coll. Publications, vol. XXVII, New York, 1942).

---

[4] Instead of that condition, one may only suppose that the closed subrings considered are right vector spaces over K, and that for every $u \neq 0$ in K, there exists $\xi \varepsilon T$ such that $\xi(u) \neq 0$; both conditions imply that T contains the subsfield of $\mathcal{E}$ consisting of all right multiplications $x \rightarrow xu$ by elements of K (which may be identified with K), and conversely they are implied by that assumption.

LOCALLY COMPACT RINGS. II.*

By Irving Kaplansky.**

The purpose of this note is to complete the theory of the radical of a locally compact ring that was initiated in [4], and to prove two decomposition theorems illustrating the role of local direct sums in the theory.

Theorem 1. *The radical of a locally compact ring is closed; more precisely, it is the intersection of the closed regular maximal right (or left) ideals.*

Proof. Let A be a locally compact ring, R its radical, R' the closure of the latter. Then R' is again a locally compact ring, and by [4, Lemma 2] its radical is R. If $R \neq R'$ then [4, Th. 11] R' contains a closed regular maximal right ideal M. Then $R \subset M$, which contradicts the fact that R is dense in R'.

We turn to the second statement of the theorem. Now that we know that the radical is closed, it will evidently suffice to treat the semi-simple case. Then by applying [4, Th. 2] we can reduce further to the case where A is totally disconnected. Let B be a compact open subring [4, Lemma 4], S the radical of B, and e an idempotent in B mapping into the unit element of B/S [3, Th. 16 and Lemma 12]. We shall now prove the following statement: any regular maximal right ideal M in A containing eA is closed. Suppose on the contrary that M is dense in A, and let f be a left unit element modulo M. Then M contains elements $f + x$ with x arbitrarily small, say so small that $r = x - ex$ is in S and is consequently quasi-regular: $r + s + rs = 0$. Now $ex \in M$ and so $f + r \in M$, $fs + rs \in M$. Since f is a left unit modulo M, $fs - s \in M$. Hence $s + rs \in M$, $r \in M$, and $f \in M$, a contradiction.

Let now y be an element lying in every closed regular maximal right ideal of A. For every idempotent g of finite rank in B, we have $gy = 0$ by [4, Lemma 8]. Since e is a limit of such idempotents, $ey = 0$. Consider the ideal $I = eA + (1 + y - ye)A$ (this is well-defined even if there is no

---

* Received May 21, 1949.

** This paper was written while the author was a fellow of the John Simon Guggenheim Memorial Foundation, on leave from the University of Chicago.

unit element). If $I \neq A$, it can be expanded to a right ideal M maximal with respect to exclusion of $y - ye$. Then M will be actually maximal, regular since $ye - y$ is a left unit, and closed by the preceding paragraph. This means that y is in M, a contradiction. It must therefore be the case that I is all of A. Write

$$eb + (1 + y - ye)c = ye - e - y.$$

After left-multiplication by $1 - e$ this becomes $(y - ye) o (c - ec) = 0$. Thus $y - ye = y(1 - e)$ is right quasi-regular and so is $(1 - e)y = y$, as follows from the last identity on p. 154 of [3]—the lack of a unit element does not matter. Hence the intersection of the closed regular maximal right ideals consists entirely of right quasi-regular elements, and is therefore 0. Of course, by symmetry the same is true for left ideals.

An immediate corollary of Theorem 1 is the fact that a locally compact semi-simple ring is a subdirect sum of locally compact primitive rings. In the study of the kind of subdirect sum that may arise, it seems certain that a leading role will be played by the local direct sum introduced by Braconnier [2].

Definition. Let A_i be topological rings with open subrings B_i. The *local direct sum* A of A_i relative to B_i is the set of $\{a_i\}$, where $a_i \varepsilon A_i$ and all but a finite number are in B_i; the subring where all coordinates are in B_i is given the Cartesian product topology and its neighborhoods of 0 are declared to be neighborhoods of 0 in A.

It is readily verified that this definition makes A a topological ring; moreover A is locally compact if and only if each A_i is locally compact and all but a finite number of B's are compact.

The next theorem shows that in a slight generalization of the commutative case a complete reduction to Q-rings is possible.

THEOREM 2. *A locally compact ring whose idempotents are in the center is a local direct sum of Q-rings.*

Proof. We begin by reducing the problem to the totally disconnected case. Let A be the given ring, C its component of 0. In C we know [4, p. 448] that there is an ideal P such that $P^2 = 0$ and C/P is an algebra of finite order over the real numbers. The latter is the direct sum of an algebra with unit element f and a nilpotent algebra. Let $g \varepsilon C$ be an idempotent mapping on f modulo P. We take the Peirce decomposition

160,850.

of A relative to g, and we may confine further study to the component $B = A(1 - g)$. In B the component D of 0 is nilpotent. Let E be a compact open subring of the totally disconnected ring B/D, and choose an idempotent h mapping on the unit element of E modulo its radical. We further pick an idempotent e in B mapping on h modulo D. In the Peirce decomposition of B relative to e, the component $B(1 - e)$ is a Q-ring. It remains to study the totally disconnected ring Be.

Let us start with A totally disconnected, F a compact open subring, and e, e_i idempotents mapping respectively on the unit element and the primitive idempotents of F modulo its radical. After a Peirce decomposition we may assume that e is the unit element of A. We consider the mapping $a \rightarrow \{ae_i\}$, which is at least an algebraic isomorphism. The compact ring F is the full Cartesian direct sum of $\{Fe_i\}$, as follows from [3, Th. 17]. Moreover a sequence with all but a finite number of its coordinates in Fe_i will arise from a suitable element of A. Thus at least the local direct sum occurs (in its correct topology). Our theorem will be proved as soon as we verify that for each $a \varepsilon A$, all but a finite number of ae_i are in Fe_i. There is a neighborhood U of 0 such that $aU \subset F$. We may suppose that $U \subset F$, and such a neighborhood of F consists of elements restricted at only a finite number of coordinates. At all the remaining coordinates ae_i must be in Fe_i. This completes the proof of Theorem 2.

Let us now consider a locally compact totally disconnected Q-ring A. Then any compact open subring of A has an open radical. In a context where all subrings of A are semi-simple we can conclude that A is discrete. All subrings of A are semi-simple for example if A is a Boolean ring, or more generally if every element satisfies $a^{n(a)} = a, n(a) > 1$. Combining this remark with Theorem 2, we see that a locally compact ring satisfying $a^{n(a)} = a$ is a local direct sum of discrete rings. We shall now prove a more general result by classifying all strongly regular locally compact rings. As defined in [1], a ring is strongly regular if for every a there exists an x with $a^2x = a$; we shall make free use of the results in [1].

THEOREM 3. *A strongly regular locally compact ring A is of the form $X \oplus Y \oplus Z$ where X is discrete, Y is the direct sum of a finite number of non-discrete division rings, and Z is a local direct sum of discrete strongly regular rings relative to finite subfields.*

Proof. By [4, Th. 2] we can confine ourselves to the totally disconnected case, and we first treat the case where A is a Q-ring. Then A

contains a compact open radical subring B. Let M be an open maximal ideal in A, so that A/M is a discrete division ring. Then B maps into a finite radical subring of A/M; such a subring is a nilring and hence 0. Thus B is contained in M. If we denote by I the intersection of all open maximal ideals, then I is open since it contains B. We proceed to examine I, which, as a ring on its own merits is again strongly regular. Its maximal ideals are in one-to-one correspondence with the non-open maximal ideals of A. Thus for every maximal ideal N in I, I/N is a non-discrete locally compact division ring. We shall now prove that there are only a finite number of N's. If we suppose the contrary, then I has an infinite set of orthogonal idempotents e_i (this follows from the results in [1], or it can be proved by a simple direct argument). For each i we can find an element b_i in B such that e_ib_i is a non-zero element of B. Write $c_n = e_1b_1 + \cdots + e_nb_n$, $a_n = e_1b_1 + e_2b_2^2 + \cdots + e_nb_n^n$ and let $c, a \in B$ be limit points of the sequences c_n, a_n respectively. Find the element x with $a^2x = a$. Now $c^r \to 0$ since B is topologically nilpotent; hence $xc^r \to 0$, $xc^r \in B$ for a suitably large r. Take any $s > r$, and examine the situation modulo a maximal ideal N in I not containing e_sb_s. The division ring I/N has a unique valuation associated with it which we shall denote by V, and we may apply V in a natural way to all elements of I. Since e_sb_s is topologically nilpotent, $V(b_s) > 0$. Next we note $e_sc_n = e_sb_s$ for all $n \geqq s$ and hence $e_sc = e_sb_s$; similarly $e_sa = e_sb_s^s$. From $a^2x = a$ we reduce $V(x) = -sV(b_s)$. On the other hand, from the fact that xc^r is in B we deduce $V(x) + rV(b_s) > 0$, a contradiction.

At this point we have proved that I is the direct sum of a finite number of non-discrete division rings. Thus it has a unit element and is a direct summand of A; moreover A is the direct sum of I and the discrete ring A/I. This concludes the discussion of the case where A is a Q-ring.

In the general case we apply Theorem 2 (applicable since the idempotents are in the center) to learn that A is a local direct sum of Q-rings. Each summand is strongly regular and its structure is known from the preceding discussion. From this we can see that all but a finite number of summands must be discrete—otherwise the local sum would not be strongly regular. In the local sum over the discrete summands, all but a finite number of the subrings (relative to which the local sum is taken) must be finite. Those summands for which the subring in question is 0 may be lumped together with the finite number of exceptions already encountered to form a discrete summand. The remaining local sum is taken relative to finite non-zero subrings all but a finite number of which must be strongly regular. Since a finite strongly regular ring is a direct sum of finite fields, the sum may be further decomposed so as to be taken to finite fields (the finite number of exceptions are again moved over to the discrete summand).

INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

[1] R. Arens and I. Kaplansky, "Topological representation of algebras," *Transactions of the American Mathematical Society*, vol. 63 (1948), pp. 457-481.

[2] J. Braconnier, "Sur les groupes topologiques localement compacts," *Journal de Mathématiques Pures et Appliquées*, vol. 27 (1948), pp. 1-85.

[3] I. Kaplansky, "Topological rings," *American Journal of Mathematics*, vol. 69 (1947), pp. 153-183.

[4] ———, "Locally compact rings," *American Journal of Mathematics*, vol. 70 (1948), pp. 447-459.

PRIME IDEALS AND THE LOWER RADICAL.*

By JAKOB LEVITZKI.

The lower radical L of a ring R has been defined by Baer [1][1] as the intersection of all the radical ideals [2] of the ring. Baer has shown that L itself is a radical ideal. The lower radical of an ideal [3] A may be defined as the ideal $L(A)$ which is uniquely determined by the following two properties: 1) $A \subseteq L(A)$. 2) $L(A)/A$ is the lower radical of R/A. Denote by $M(A)$ the radical of A in the sense of McCoy [4], i. e. the set of all elements r with the property that every m-system [4] which contains r contains an element of A. McCoy has shown that $M(A)$ is an ideal, $M(A)$ is the intersection of all the prime ideals which contain A and finally that $M(A)/A$ is a radical ideal in R/A. The radical M of the ring is defined as the radical of the zero ideal. McCoy remarks that his radical $M(A)$ "is contained in Fitting's radical, but the exact relation between these concepts is an unsolved problem." He also says that "the relation of M to the radicals of Koethe and Levitzki is an unsolved problem." In the present note we solve these problems by showing that *the radical ideal $M(A)/A$ of R/A is actually the lower radical of R/A, i. e. $M(A) = L(A)$.* In particular $M = L$. This implies that M is semi-nilpotent and that M is distinct from the maximal semi-nilpotent ideal of the ring, and hence also from the radical in the sense of Koethe (see [1], [2], and [3]).[5] The same applies to the radical of an ideal. On the other hand McCoy's results combined with our theorem yield the following characterization of the lower radical: *The lower radical of a ring is the intersection of all the prime ideals of the ring.* In the following we use the concept of an m-sequence, which is actually a special case of an m-

---

* Received January 11, 1950.

[1] Numbers in brackets [ ] refer to the bibliography at the end of the paper.

[2] An ideal A is a radical ideal if A is a nil-ideal and R/A has no nil-potent ideals other than zero.

[3] Ideal = two sided ideal.

[4] A set S is an m-system if $a \in S$, $b \in S$ implies the existence of $c \in R$ such that $acb \in S$.

[5] In an example given by Baer ([1], § 2) we have $R \neq 0$, $L = 0$ while the radical in the sense of Koethe etc. is $= R$.

system as defined by McCoy. Incidentally this concept yields yet another characterization of the lower radical.[6]

We begin with the following definitions:

Definition 1. An infinite sequence of elements in a ring R

$$a_0, a_1, a_2, \cdots \tag{1}$$

is an m-sequence if for each non-negative integer n there exists an element b_n in R so that $a_{n+1} = a_n b_n a_n$.

Definition 2. An m-sequence is a vanishing m-sequence if some term of the sequence is $= 0$.

Remark. If some term of an m-sequence is $= 0$, then evidently also all subsequent terms are $= 0$. Hence a vanishing m-sequence contains at most a finite number of nonzero terms. Moreover, since each term of an m-sequence is a left factor as well as a right factor of all subsequent terms, we have the following

LEMMA 1. *If some term of an m-sequence is contained in a right (left) ideal A, then also all subsequent terms are contained in A.*

We need also the following

LEMMA 2. *The elements of an m-sequence form an m-system.*

Proof. Consider an m-sequence (1). We have to show that for any pair of terms a_i, a_k of (1) there exists a term a_j and an element b_{ik} of R such that $a_j = a_i b_{ik} a_k$. Now it is an easy consequence of the defining relation $a_{n+1} = a_n b_n a_n$ that this holds for any a_j with $j > i$, $j > k$.

Definition 3. An element a of a ring R and an infinite sequence $b_0, b_1, \cdots$ are associated with each other if $b_0 = a$ and for some infinite sequence $c_0, c_1, \cdots$ we have $b_{n+1} = b_n c_n b_n$, $i = 0, 1, 2, \cdots$.

As an immediate consequence of definitions 1 and 3 we obtain

LEMMA 3. *Each infinite sequence which is associated with some element of a ring R is an m-sequence.*

Remark. Conversely, each m-sequence (1) is evidently associated with a_0. Hence Lema 3 gives a method for constructing all possible m-sequences.

---

[6] As will be shown elsewhere a certain generalization of this concept may be used with advantage also in other problems concerning nil-rings.

THEOREM 1. *An element a of a ring R belongs to the lower radical L of R if and only if each sequence which is associated with a is a vanishing m-sequence.*

Proof. For the proof of the "only if" part of the theorem we need the following characterization of the lower radical: Denote by N_1 the ordinary radical of R, i. e. the sum of all nilpotent ideals of R. If $\alpha > 1$ is an ordinal such that $\alpha = \beta + 1$, then the ideal N_α is defined by the relation $N_\beta \subseteq N_\alpha$ and the requirement that N_α / N_β should be the ordinary radical of R/N_β. If α is a limit ordinal, then N_α is defined as the join of all N_γ with $\gamma < \alpha$. For some ordinal τ of minimal value we have then $N_{\tau+1} = N_\tau$ and it follows that $L = N_\tau$ (see [1], page 539). Suppose now that $a \varepsilon L$ and consider a sequence (1) which is associated with a, so that we have $a_0 = a$. By Lemma 1 it follows that all terms of (1) are elements of L. Hence each element of (1) is contained in some N_γ. Thus for each non-negative integer i there exists a smallest ordinal α_i so that $a_i \varepsilon N\alpha_i$. If for some i the ordinal α_i were a limit ordinal, it would follow from the definition of $N\alpha_i$ as the join of all N_γ with $\gamma < \alpha_i$, that $a_i \varepsilon N_\gamma$ for some $\gamma < \alpha_i$, which is a contradiction to the minimum property of α_i. Hence none of the ordinals α_i is a limit ordinal. Denote by k a non-negative integer so that $\alpha_k =$ minimum α_i, $i = 0, 1, 2, \cdots$. Since α_k is not a limit ordinal we have either $\alpha_k = 1$ or $\alpha_k = \beta + 1$ with $\beta \geq 1$. Suppose first that we are confronted by the second possibility, then by the definition of $N_{\beta+1}$ it follows that for some positive integer m the right ideal $(a_k, a_k R)$ generated by a_k satisfies the relation:

$$(2) \qquad (a_k, a_k R)^m \subseteq N_\beta.$$

Now we have $a_{k+1} = a_k b_k a_k \varepsilon (a_k, a_k R)^2$, and for an arbitrary positive integer p it follows easily that $a_{k+p} \varepsilon (a_k, a_k R)^{2^p}$. Hence by choosing p so that $2^p \geq m$ we obtain

$$(3) \qquad a_{k+p} \varepsilon (a_k, a_k R)^{2^p} \subseteq (a_k, a_k R)^m \subseteq N_\beta.$$

Relation (3) shows that $\alpha_{k+p} \leq \beta < \alpha_k$, which is a contradiction to the minimum property of α_k. This contradiction implies that we must have $\alpha_k = 1$ and thus relations (2) and (3) may be replaced by $(a_k, a_k R)^m = 0$ and $a_{k+p} = 0$ respectively, which shows that each sequence which is associated with a is a vanishing m-sequence. This concludes the proof of the "only if" part of the theorem. The "if" part will be settled if we can show that any element a such that $a \notin L$ is associated with a non-vanishing m-sequence. We begin the construction of such a sequence by setting $a_0 = a$. Since L is a radical ideal, i. e., R/L has no nilpotent ideals, it follows by $a_0 \not\varepsilon L$ that also $a_0R \cdot a_0R \not\varepsilon L$. Hence for some $b_0 \varepsilon R$ we have $a_0b_0a_0 \not\varepsilon L$. We define the second term of the required sequence by putting $a_1 = a_0b_0a_0$. Since $a_1 \not\varepsilon L$ we may repeat the above procedure with a_1 and thus obtain the third term $a_2 = a_1b_1a_1 \not\varepsilon L$. By continuing this process we obtain (by induction) a non-vanishing m-sequence which is associated with a. This completes the proof of the theorem.

Theorem 2. *The radical of a ring R in the sense of McCoy coincides with the lower radical L.*

Proof. Denote by M the radical in the sense of McCoy and suppose first that for some element a we have $a \varepsilon M$, $a \not\varepsilon L$. By the "if" part of Theorem 1 it follows that there exists a non-vanishing m-sequence which is associated with a. By Lemmas 2 and 3 we known that the elements of such a sequence form an m-system. Since this system contains a and does not contain 0 we have a contradiction to $a \varepsilon M$. This shows that if $a \varepsilon M$ then also $a \varepsilon L$, i. e. $M \subseteq L$. Suppose now[7] that there exists an element a such that $a \varepsilon L$, $a \not\varepsilon M$. By the definition of M it follows that there exists an m-system S which contains a and does not contain 0. Hence for some $b_0 \varepsilon R$ we have $0 \neq ab_0a \varepsilon S$. Put $a_0 = a$ and $a_1 = a_0b_0a_0$. Since $a_1 \varepsilon S$ we may repeat the above procedure with a_1, and thus by continuing the process obtain (by induction) a non-vanishing m-sequence which is associated with a. Since by Theorem 1 the existence of such a sequence is incompatible with $a \varepsilon L$ it follows that we must have $a \varepsilon M$, or $L \subseteq M$ which in conjunction with $M \subseteq L$ yields the required result.

Corollary 1. *For an ideal A of a ring R we have $M(A) = L(A)$.*

By combining Theorem 2 with McCoy's results [4] we obtain

Corollary 2. *The lower radical of a ring is the intersection of all the prime ideals of the ring.*

Since the lower radical of a ring is $= 0$ if and only if the zero ideal is the only nilpotent ideal of the ring we have by [4], Theorem 6:

Corollary 3. *A necessary and sufficient condition that a ring be iso-*

---

[7] The proof may be completed as follows: Since M is a radical ideal (see [4]) we have $M \supseteq L$, which in conjunction with $L \supseteq M$ yields $M = L$. Now the relation $M \supseteq L$ has been proved by McCoy by using the concept and the properties of a prime ideal. The proof of $M \supseteq L$ in the text does not involve prime ideals and uses the "only if" part of Theorem 1 instead.

morphic to a subdirect sum of prime rings is that the given ring does not contain nilpotent ideals other than zero.

Finally we have

COROLLARY 4. *The intersection of all the prime ideals of a ring is* $= 0$ *if and only if the zero ideal is the only nilpotent ideal of the ring.*

JERUSALEM UNIVERSITY.

REFERENCES.

[1] R. Baer, "Radical ideals," *American Journal of Mathematics*, vol. 65 (1943), pp. 537-568.

[2] J. Levitzki, "On the radical of a general ring," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 462-466.

[3] ———, "On three problems concerning nil-rings," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 913-919.

[4] N. H. McCoy, "Prime ideals in general rings," *American Journal of Mathematics*, vol. 71 (1949), pp. 823-833.

BANACH ALGEBRAS OF CONTINUOUS FUNCTIONS.*

By Bertram Yood.

1. **Introduction.** For the Banach algebra $C(X)$ of all bounded, continuous and complex-valued functions defined on a completely regular space with the usual definition of the norm, the space of maximal ideals is homeomorphic to the Čech β-compactification of X [6]. In this paper we investigate the structure of the space of maximal ideals of the more general Banach algebras $C(X, A)$ of all bounded continuous functions defined on a space X as above with values in a commutative complex Banach algebra A with a unit.

In § **3** it is shown that if A is also a B^*-algebra (in the sense of [11]) then the space of maximal ideals of $C(X, A)$ is homeomorphic to the β-compactification of the Cartesian product space $X \times \mathfrak{M}$, where $\mathfrak{M}$ is the space of maximal ideals of A. This situation is further investigated in § **4** under the additional assumptions that X and $X \times \mathfrak{M}$ are normal. This yields other descriptions of the space of maximal ideals of $C(X, A)$ and hence alternative descriptions of the β-compactification of $X \times \mathfrak{M}$. This is given in terms of the β-compactification of X, the Cartesian product space $P_{t \epsilon X}\mathfrak{M}$ and continuous functions on $X \times \mathfrak{M}$ and closure in $X \times \mathfrak{M}$.

In § **5** it is assumed that X is compact ($=$ bicompact) but the requirement that A be a B^*-algebra is dropped. Here it is shown that $C(X, A)$ may have X as its space of maximal ideals without A being the algebra of scalars, specifically for A a primary ring (i. e., a ring with one maximal ideal).

An investigation of rings $C(X, A)$ of all continuous functions on a topological space X to a topological ring A was made by Kaplansky [10]. Where the hypotheses are comparable, the results of Kaplansky hold for much more general rings A but much greater demands are made on the spaces X than in this paper. Kaplansky also considers rings of all continuous functions (unbounded as well as bounded). Here we use only bounded functions and endeavor to exploit as fully as possible the techniques of commutative Banach algebra theory as developed by Gelfand and others.

A thorough exposition of the theory of the algebra $C(X)$ (for the real case where the theory is quite the same) can be found in a paper of Hewitt

---

* Received March 3, 1950; presented to the American Mathematical Society, April 1, 1949.

[7]. Hewitt also investigates the algebra of all (bounded and unbounded) continuous real-valued functions defined on the completely regular space X.

2. Definitions and preliminaries.

We use the letter A to represent a complex Banach algebra (normed ring) which is commutative and posseses a unit e of norm one. For the fundamental properties of A see [3] or [9]. Associated with A is its space $\mathfrak{M}$ of maximal ideals topologized as in [3] to become a compact Hausdorff sapce. Each $M \varepsilon \mathfrak{M}$ determines a homomorphism $X(M)$ of A onto the complex numbers K. M consists of all $x \varepsilon A$ for which $x(M) = 0$ and for each x, $x(M)$ is a continuous function of M. As a function of x for a fixed M, $x(M)$ is a multiplicative linear functional defined on A and all such functionals are of this form.

For a topological space X, let $C(X, A)$ be the collection of all bounded continuous functions defined on X with values in A. With the norm $\| f(\cdot) \| = \sup \| f(t) \|$, $t \varepsilon X$, $C(X, A)$ is a commutative Banach algebra with a unit.

2.1. Lemma. *Let t_0 be a point of X and M a maximal ideal of A. Then the set N of all $f(\cdot) \varepsilon C(X, A)$ for which $f(t_0) \varepsilon M$ forms a maximal ideal of $C(X, A)$.*

Suppose that $g(\cdot) \varepsilon C(X, A)$ and $g(t_0) \not\varepsilon M$. Then there exist elements $v \varepsilon A$, $w \varepsilon M$ such that $vg(t_0) + w = e$. We let $h(\cdot) \varepsilon C(X, A)$ be the function with the constant value v and $e(\cdot)$ the function with the constant value e. Then

$$e(\cdot) = h(\cdot)g(\cdot) + \{e(\cdot) - h(\cdot)g(\cdot)\}$$

where $e(\cdot) - h(\cdot)g(\cdot) \varepsilon N$ since $e - vg(t_0) = w$. This shows that $g(\cdot)$ and N generate the ideal $C(X, A)$ and therefore that N is maximal.

On the basis of this result we divide the ideals into two classes which are the analogues of those used by Hewitt [7]. An ideal I in $C(X, A)$ is called *fixed* if there exist $t_0 \varepsilon X$ and $M \varepsilon \mathfrak{M}$ such that

$$I \subset \{f(\cdot) \varepsilon C(X, A) \mid f(t_0) \varepsilon M\}.$$

Otherwise the ideal I is called *free.*

For $f(\cdot) \varepsilon C(X, A)$, $t \varepsilon X$, $M \varepsilon \mathfrak{M}$, we write $f(t)(M)$ as $f(t, M)$.

2.2. Lemma. *For each $f(\cdot) \varepsilon C(X, A)$, $f(t, M)$ is a continuous function on the Cartesian product space $X \times \mathfrak{M}$.*

Let (t_0, M_0) be a point of $X \times \mathfrak{M}$ and $\epsilon > 0$. Now

$$| f(t, M) - f(t_0, M_0) | \leqq | f(t, M) - f(t_0, M) | + f(t_0, M) - f(t_0, M_0) |$$

and $|f(t,M)-f(t_0,M)| \leqq \| f(t)-f(t_0)\|$. We choose a neighborhood $U(t_0)$ of t_0 in X for which $\| f(t)-f(t_0)\| < \epsilon/2$, and a neighborhood $V(M_0)$ of M_0 in $\mathfrak{M}$ for which $|f(t_0,M)-f(t_0,M_0)| < \epsilon/2$. For (t,M) in $U(t_0)\times V(M_0)$ we have $|f(t,M)-f(t_0,M_0)| < \epsilon$.

Following Rickart [11] we call a Banach algebra a "B^*-algebra" if to each element x there corresponds a unique element x', called the adjoint of x where (i) $(x')' = x$, (ii) $(xy)' = y'x'$, (iii) for α and β complex numbers and $\bar{\alpha}$ and $\bar{\beta}$ their complex conjugates, $(\alpha x + \beta y)' = \bar{\alpha}x' + \bar{\beta}y'$, (iv) $\| x'x \| = \| x \|^2$.

The B^*-algebras considered in this paper are commutative and possess a unit of norm one. It is known that $\| x' \| = \| x \|$ and that such a B^*-algebra A is equivalent to the algebra of all complex-valued continuous functions defined on its space of maximal ideals $\mathfrak{M}$ ([5, Lemma 1]). Also $\| x \| = \sup |x(M)|$, $M \in \mathfrak{M}$, $x'(M) = \overline{x(M)}$ for each $M \in \mathfrak{M}$, and A is without radical. As the topology for $\mathfrak{M}$, we use that of [3] in which the neighborhoods of $M_0 \in \mathfrak{M}$ are all sets of the form

$$\{M \in \mathfrak{M} \mid |x_i(M) - x_i(M_0)| < \epsilon,\ i = 1,\cdots,n\}$$

where $x_1,\cdots,x_n$ are elements of A and $\epsilon > 0$. This yields a compact Hausdorff topology for $\mathfrak{M}$. Two other topologies are investigated by Gelfand and Šilov in [6]. The first one, introduced by Stone [15], may be described as follows. For each set $\mathfrak{A} \subset \mathfrak{M}$, $M_0 \in \bar{\mathfrak{A}}$ if and only if $M_0 \supset \bigcap M$, $M \in \mathfrak{A}$. The other topology can be employed in any commutative Banach algebra A with a unit for which given $x \in A$ there is $y \in A$ with $x(M) = \overline{y(M)}$, for every $M \in \mathfrak{M}$. Here $M_0 \in \bar{\mathfrak{A}}$ if and only if $M_0 \subset \overline{\bigcup M}$, the union taken over $M \in \mathfrak{A}$. Equivalently $M_0 \in \bar{\mathfrak{A}}$ if and only if for every $x \in M_0$, $\inf |x(M)| = 0$, $M \in \mathfrak{A}$. This topology can be employed for our B^*-algebras. In view of results in [6], the theorem of Gelfand [3] which states that any compact Hausdorff topology for $\mathfrak{M}$ in which all the functions $x(M)$ are continuous is homeomorphic to $\mathfrak{M}$ in the first topology above and the fact that a B^*-algebra is equivalent to $C(\mathfrak{M})$ ($\mathfrak{M}$ with the first topology), all three topologies for $\mathfrak{M}$ yield homeomorphic spaces.

For such a B^*-algebra A, $C(X,A)$ is also a B^*-algebra to which the above remarks about its maximal ideals apply.

3. X completely regular, A a B^*-algebra. We obtain the structure of the space of maximal ideals of $C(X,A)$ in this case as the β-compactification in the sense of Čech [2] of the completely regular space $X \times \mathfrak{M}$. The compactification procedure of Čech is equivalent to the compactification procedure of Gelfand and Šilov [6].

3.1. THEOREM. *Let X be a completely regular space, A a commutative B^*-algebra with a unit. Then the space of maximal ideals of $C(X, A)$ is homeomorphic to the β-compactification of $X \times \mathfrak{M}$.*

In Lemma 2.2 it was shown that for any $f(\cdot) \varepsilon C(X, A)$, $f(t, M)$ is continuous on $X \times \mathfrak{M}$. Conversely we show that if $f(t, M)$ is continuous on $X \times \mathfrak{M}$, there exists $g(\cdot) \varepsilon C(X, A)$ such that $f(t, M) = g(t, M)$ for each $t \varepsilon X$, $M \varepsilon \mathfrak{M}$. Let $t_0 \varepsilon X$. For each $M' \varepsilon \mathfrak{M}$, there exists a neighborhood $V(M')$ of M' and a neighborhood $U(M', t_0)$ of t_0 in X such that for all $t \varepsilon U(M', t_0)$, $M \varepsilon V(M')$ we have $| f(t, M) - f(t_0, M') | < \epsilon$. A finite number $V(M_1), \cdots, V(M_n)$ of the sets $V(M')$ cover $\mathfrak{M}$. Let $U(t_0) = \bigcap U(M_i, t_0)$. Then for each $M \varepsilon \mathfrak{M}$, $| f(t, M) - f(t_0, M) | < \epsilon$ for $t \varepsilon U(t_0)$.

Next the function $g(\cdot)$ is defined for $t = t_0$. $g(t_0)$ is taken to be that element in A such that for each $M \varepsilon \mathfrak{M}$, $g(t_0)(M) = f(t_0, M)$. Since $f(t_0, M)$ is continuous in M and A is a B^*-algebra, this is possible. This procedure is followed for each element in X. Also $g(\cdot) \varepsilon C(X, A)$. For given $\epsilon > 0$ there is a neighborhood $U(t_0)$ such that for $t \varepsilon U(t_0)$,

$$\| g(t) - g(t_0) \| = \sup_{M \varepsilon \mathfrak{M}} | g(t)(M) - g(t_0)(M) | = \sup | f(t, M) - f(t_0, M) | \leqq \epsilon$$

by the above analysis. ($f(t, M)$ is assumed bounded.)

The correspondence $f(\cdot) \rightarrow f(t, M)$ between $C(X, A)$ and $C(X \times \mathfrak{M})$, the Banach algebra of all complex-valued, bounded, continuous functions defined on $X \times \mathfrak{M}$, is next considered. It is easy to verify that this is a 1-1 correspondence of $C(X, A)$ onto $C(X \times \mathfrak{M})$ which preserves all the algebraic operations. Furthermore since

$$\| f(\cdot) \| = \sup_{t \varepsilon X} \| f(t) \| = \sup_{t \varepsilon X} \sup_{M \varepsilon \mathfrak{M}} | f(t, M) |$$

and since

$$\sup_{M \varepsilon \mathfrak{M}} | f(t, M) | = | f(t, M_t) |$$

for a suitably chosen $M_t \varepsilon \mathfrak{M}$,

$$\| f(\cdot) \| = \sup_{t \varepsilon X} | f(t, M_t) | = \sup_{(t, M) \varepsilon X \times \mathfrak{M}} | f(t, M) | ,$$

the correspondence is an equivalence. Now $X \times \mathfrak{M}$ is completely regular and hence the space of maximal ideals of $C(X \times \mathfrak{M})$ is homeomorphic to the β-compactification of $X \times \mathfrak{M}$ (see [6]). This is then the case for the space of maximal ideals of $C(X, A)$.

4. X and $X \times \mathfrak{M}$ normal, A a B^*-algebra. Under these conditions we give other descriptions of the space of maximal ideals of $C(X, A)$. By

Theorem 3.1, this then gives us descriptions of the β-compactification of $X \times \mathfrak{M}$. In this we use the points of the Wallman compactification of X (which is equivalent to the β-compactification since X is normal), the Cartesian product space $P_{t\varepsilon X}\mathfrak{M}$ and the continuous functions on $X \times \mathfrak{M}$. The points of the compactification are pairs of elements of the compactification of X and equivalence in the Cartesian product space.

We wish to emphasize the fact that the more complicated description of the space of maximal ideals of $C(X, A)$ which is given in this section has the virtue (not possessed by the analysis of § **3**) that it yields the general form of a multiplicative linear functional defined on $C(X, A)$.

For $f(\cdot) \varepsilon C(X, A)$ we define the set

$$(4.1) \qquad X(f, \epsilon) = \{t \varepsilon X \mid |f(t, M)| \leqq \epsilon \text{ for at least one } M \varepsilon \mathfrak{M}\}.$$

Let I be a non-trivial ideal of $C(X, A)$. We denote by $\mathfrak{A}(I)$ the collection of all sets of the form $X(f, \epsilon)$, $f \varepsilon I$, for a fixed $\epsilon > 0$. Inasmuch as I is a linear manifold, $\mathfrak{A}(I)$ is independent of the choice of $\epsilon > 0$ in its definition.

4.1. LEMMA. *The family of sets* $\mathfrak{A}(I)$ *has the finite intersection property.*

Let $B(X, A)$ be the B^*-algebra of all bounded functions defined on X with range in A where $\| f(\cdot) \| = \sup \|f(t)\|$, $t \varepsilon X$. Then $C(X, A)$ is a B^*-subalgebra of $B(X, A)$. I is contained in a maximal ideal N of $C(X, A)$. By a theorem of Šilov [12, p. 84], N is contained in a maximal ideal N_1 of $B(X, A)$. If we define the sets $X(f, \epsilon)$ for each $f \varepsilon N_1$ then the collection so obtained has the finite intersection property [18, Lemma 2.2]. Since $\mathfrak{A}(I)$ is a subcollection, it shares this property.

4.2. LEMMA. $\mathfrak{A}(I)$ *is contained in a maximal additive ideal in the distributive lattice of all closed subsets of* X.

We show first that the sets $X(f, \epsilon)$ are closed. Let t_0 be in the closure of $X(f, \epsilon)$. Suppose that $\inf |f(t_0, M)| = \eta > \epsilon$, $M \varepsilon \mathfrak{M}$. Then there is a neighborhood $U(t_0)$ of t_0 such that for each $t \varepsilon U(t_0)$, $\| f(t) - f(t_0) \| < (\eta - \epsilon)$. For each such t and each $M \varepsilon \mathfrak{M}$ we have

$$|f(t, M)| \geqq |f(t_0, M)| - |f(t_0, M) - f(t, M)| > \epsilon$$

inasmuch as $|f(t_0, M) - f(t, M)| \leqq \| f(t_0) - - f(t) \|$. This is a contradiction. Since $f(t_0, M)$ is a continuous function on the compact set $\mathfrak{M}$, there is an $M \varepsilon \mathfrak{M}$ for which $|f(t_0, M)| \leqq \epsilon$ and $X(f, \epsilon)$ is closed.

Thus $\mathfrak{A}(I)$ having the finite intersection property is contained in a maximal additive ideal in the distributive lattice of closed subsets of X; see [17, p. 114]. We use the letter $\mathfrak{L}$ to designate this lattice. By an *additive ideal* $\mathfrak{B}$ *in* $\mathfrak{L}$ is meant a collection which has the properties that if $F \varepsilon \mathfrak{B}$, $G \varepsilon \mathfrak{L}$, then $F \cup G \varepsilon \mathfrak{B}$ and if $F \varepsilon \mathfrak{B}$, $G \varepsilon \mathfrak{B}$ then $F \cap G \varepsilon \mathfrak{B}$. A *maximal additive ideal* (m. a. ideal) is one which is contained in no proper additive ideal of $\mathfrak{L}$.

We designate by $\mathfrak{S}$, the Boolean algebra of subsets of X which is generated by the open sets and the closed set of X. See [1] for the lattice-theoretic terminology employed. The following lemma on distributive lattices is needed.

4. 3. LEMMA. *Let* $\mathfrak{A}$ *be a lattice with a zero element* Λ. *Let* $\mathfrak{B}$ *be a maximal additive ideal in* $\mathfrak{A}$. *A real valued function* μ *defined on* $\mathfrak{A}$ *for which* $\mu(E) = 1$, $E \varepsilon \mathfrak{B}$, $\mu(E) = 0$, $E \not\varepsilon \mathfrak{B}$ *is additive in the sense that if* $E_1 \cap E_2 = \Lambda$, *then* $\mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2)$.

It is to be assumed that $\mathfrak{A}$ is not the trivial lattice consisting of just its zero element. Wallman [7] has shown that for $\mathfrak{B}$ to be a m. a. ideal it is necessary and sufficient that $\mathfrak{B}$ be a maximal collection of elements with the finite intersection property, and that this is equivalent to the possession by $\mathfrak{B}$ of the properties (1) if $F, G \varepsilon \mathfrak{B}$ then $F \cap G \varepsilon \mathfrak{B}$ and (2) $F \varepsilon \mathfrak{B}$ if and only if $F \cap G \neq \Lambda$ for every $G \varepsilon \mathfrak{B}$. From this it follows readily if $E_1 \varepsilon \mathfrak{B}$ and $E_1 \cap E_2 = \Lambda$ then $E_2 \not\varepsilon \mathfrak{B}$, $\mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2)$. Also if $\mu(E_1) = \mu(E_2) = 0$ where $E_1 \cap E_2 = \Lambda$, there exist elements $F_i \varepsilon \mathfrak{B}$ such that $E_i \cap F_i = \Lambda$, $i = 1, 2$. But then $(F_1 \cap F_2) \cap (E_1 \cup E_2) = \Lambda$. However $F_1 \cap F_2 \varepsilon \mathfrak{B}$ so that $E_1 \cup E_2 \not\varepsilon \mathfrak{B}$.

The converse is true if the lattice is a Boolean algebra.

4. 4. LEMMA. *Let* $\mathfrak{A}$ *be a Boolean algebra on which there is defined a non-trivial real-valued additive function* μ *taking on the values zero and one only. Then the set* $\mathfrak{B}$ *of elements of* $\mathfrak{A}$ *for which* $\mu(E) = 1$ *forms a maximal additive ideal in* $\mathfrak{A}$.

Let $E_1, E_2 \varepsilon \mathfrak{B}$. Using the prime to denote complementation we have $E_1 = (E_1 \cap E_2) \cup (E_1 \cap E'_2)$ and $\mu(E_1) = \mu(E_1 \cap E_2) + \mu(E_1 \cap E'_2)$. Thus one and only one of the sets $E_1 \cap E_2$ and $E_1 \cap E'_2$ are in $\mathfrak{B}$. Clearly $\mu(E'_2) = 0$ so that since $E'_2 = (E_1 \cap E'_2) \cup (E'_1 \cap E'_2)$ each of the latter sets has μ-measure zero. Hence $E_1 \cap E_2 \varepsilon \mathfrak{B}$. Clearly $\Lambda \not\varepsilon \mathfrak{B}$. If $F \varepsilon \mathfrak{B}$, $G \varepsilon \mathfrak{A}$ then it is easy to see that $F \cup G \varepsilon \mathfrak{B}$. Hence $\mathfrak{B}$ is an additive ideal in $\mathfrak{A}$, hence, by Zorn's maximum principle, contained in an m. a. ideal $\mathfrak{B}_1$ of $\mathfrak{A}$. Let μ_1 be defined on $\mathfrak{A}$ to be 1 for elements on $\mathfrak{B}_1$ and zero otherwise. By Lemma 4. 3, μ_1 is additive. Also μ_1 agrees with μ on elements in $\mathfrak{B}$ and thus on the complements of such elements. Since any element is either in $\mathfrak{B}$ or is a complement of such an element, $\mu_1 \equiv \mu$ and $\mathfrak{B} = \mathfrak{B}_1$.

4.5. LEMMA. *Each maximal additive ideal* $\mathfrak{B}$ *of* $\mathfrak{L}$ *is contained in one and only one maximal addtive ideal* $\mathfrak{B}_1$ *of* $\mathfrak{S}$.

$\mathfrak{B}$ is a maximal family of closed subsets of X having the finite intersection property, as mentioned above. By Zorn's maximum principle $\mathfrak{B}$ is contained in some maximal collection $\mathfrak{B}_1$ of sets in $\mathfrak{S}$ with the finite intersection property. By Wallman [17], $\mathfrak{B}_1$ is a m. a. ideal of $\mathfrak{S}$. Suppose $\mathfrak{B}_2$ is another m. a. ideal of $\mathfrak{S}$ containing $\mathfrak{B}$. By Lemma 4. 3, the set functions ν_i defined on $\mathfrak{S}$, one for sets in $\mathfrak{B}_i$, zero otherwise are additive, $i = 1, 2$. Moreover for each i, with $E_1, E_2 \varepsilon \mathfrak{S}$, if $\nu_i(E_1) = \nu_i(E_2) = 1$, then

$\nu_i(E_1 \cup E_2) = \nu_i(E_1 \cap E_2) = 1$ since $\mathfrak{B}_i$ is an additive ideal in $\mathfrak{S}$.
For any E, $\nu_i(E') = 1 - \nu_i(E)$. If $\nu_i(E_1) = 1$, $\nu_i(E_2) = 0$, then $\nu_i(E_1 \cup E_2) = 1$. Also since $1 = \nu_i(E_1) = \nu_i(E_1 \cap E_2) + \nu_i(E_1 \cap E'_2)$ and $\nu_i(E_1 \cap E'_2) = 1$ by the preceding, we have $\nu_i(E_1 \cap E_2) = 0$. If $\nu_i(E_1) = \nu_i(E_2) = 0$, then $\nu_i(E_1 \cap E_2 = 1 - \nu_i(E'_1 \cup E'_2) = 0$, $\nu_i(E_1 \cup E_2) = 1 - \nu_i(E'_1 \cap E'_2) = 0$. Also for $F \varepsilon \mathfrak{B}$, $\nu_i(F) = 1$. For any closed set $G \not\varepsilon \mathfrak{B}$, if $\nu_i(G) = 1$, then $G \cap F \neq \Lambda$ for every $F \varepsilon \mathfrak{B}$ and thus the family $\mathfrak{B}$ may be augmented by G preserving the finite intersection property, which is impossible. Since for any open set, O, $\nu_i(O) = 1 - \nu_i(O')$, $\nu_1 = \nu_2$ for any open or closed set in X. Since $\mathfrak{S}$ is the Boolean algebra of subsets of X generated by these sets, the above rules for ν_i show that $\nu_1 = \nu_2$ for any set in $\mathfrak{S}$. Thus $\mathfrak{B}_1 = \mathfrak{B}_2$.

4. 6. LEMMA. *Let* N *be a maximal ideal of* $C(X, A)$ *and* $f_i \varepsilon N$, $i = 1, 2$. *Then there exists a function* $h \varepsilon N$ *with the property that for each* $t \varepsilon X$, $M \varepsilon \mathfrak{M}$,

$$h(t, M) = \sup_{i=1,2} | f_i(t, M) | .$$

To prove this one may proceed as in Lemma 2. 4 of [18]. With the notation used there, it is only necessary to note the function $k(\cdot)$ appearing there is continuous. This follows since

$$\| k(t_1) - k(t_2) \| = \sup_{M \varepsilon \mathfrak{M}} | \, | f(t_1, M) | - | f(t_2, M) | \, | \leqq \| f(t_1) - f(t_2) \| .$$

Clearly this result holds for any finite number of functions in N.

In the following lemma and its proof, the following notation is employed. N is a maximal ideal of $C(X, A)$ and $\mathfrak{A}(N)$ is the corresponding family of closed sets in X previously defined. Γ is the collection of all single-valued functions defined on X where for each $t \varepsilon X$, the image of t is in $\mathfrak{M}$. For $\gamma \varepsilon \Gamma$, $f(\cdot) \varepsilon C(X, A)$, $\mathfrak{A}(f, \gamma, \epsilon)$ is the set $\{t \varepsilon X \mid | f(t, \gamma(t)) | \leq \epsilon\}$. For $f(\cdot) \varepsilon N$, $\Gamma(f, \epsilon) = \{\gamma \varepsilon \Gamma \mid \mathfrak{A}(f, \gamma, \epsilon)$ contains a set in $\mathfrak{A}(N)\}$. Δ is the collection of all sets $\Gamma(f, \epsilon)$ as $f(\cdot)$ ranges over N. $\mathfrak{B}$ is a m. a. ideal in $\mathfrak{L}$ containing $\mathfrak{A}(N)$, $\mathfrak{B}_1$ is the (unique) m. a. ideal of $\mathfrak{S}$ containing $\mathfrak{B}$. μ is a function defined on $\mathfrak{S}$ to be one for elements in $\mathfrak{B}_1$, zero otherwise. By Lemmas 4.3 and 4.4, μ is additive and there is the sole additive set function which is one for sets in $\mathfrak{B}$ and zero for all other set in $\mathfrak{L}$. We say that μ is *the* $0-1$ *measure function associated with the m. a. ideal* $\mathfrak{B}$ of $\mathfrak{L}$.

The integral

$$\int_X f(t, \gamma(t))\, d\mu \tag{4.2}$$

taken in the sense of Hildebrandt [8] using the field of sets $\mathfrak{S}$ exists and is a linear functional on $C(X, A)$ (see also Smulian [14]).

4.7. LEMMA. *There exists a function* $\gamma_0 \varepsilon \Gamma$ *such that* N *is the set of zeros of* $\int_X f(t, \gamma_0(t))\, d\mu$ *for* $f \varepsilon C(X, A)$.

We argue as in [18, p. 154]. First we show that the sets in Δ have the finite intersection property. Let $\Gamma_i = \Gamma(f_i, \epsilon)$, $i = 1, 2$ be in Δ. Let $h(\cdot)$ be the function related to the $f_i(\cdot)$ by Lemma 4.6. It is easy to see that, for each $\gamma \varepsilon \Gamma$, $\mathfrak{A}(h, \gamma, \epsilon) = \bigcap \mathfrak{A}(f_i, \gamma, \epsilon)$. Since $h \varepsilon N$, there is a $\gamma \epsilon \Gamma$ such that $\mathfrak{A}(h, \gamma, \epsilon)$ contains a set in $\mathfrak{A}(N)$ (see Lemma 4.1). This set is contained in $\mathfrak{A}(f_i, \gamma, \epsilon)$ for $i = 1, 2$. Thus $\gamma \varepsilon \Gamma_1 \bigcap \Gamma_2$. Hence Δ is contained in a maximal family Δ' of subsets of Γ with the finite intersection property. Letting ν be a set function defined for all subsets of Γ to be one for sets in Δ' and zero otherwise, ν is additive. The arguments of [18, p. 155] may then be applied to give the desired results.

4.8. LEMMA. *For each maximal ideal* N *of* $C(X, A)$, $\mathfrak{A}(N)$ *is contained in one and only one maximal additive ideal of* $\mathfrak{L}$.

Let $\mathfrak{B}$ be such a m. a. ideal and μ be the associated $0-1$ measure function. If $\mathfrak{B}_1$ is another such m. a. ideal and μ_1 is its associated $0-1$ measure function then by Lemma 4.7 and the fact that two linear functionals with the same set of zeros are identical we have

$$\int_X f(t, \gamma(t))\, d\mu \equiv \int_X f(t, \gamma_1(t))\, d\mu_1 \tag{4.3}$$

for appropriate γ and γ_1 in Γ. Now since $\mathfrak{B} \neq \mathfrak{B}_1$, by Wallman's criterion for a m. a. ideal there exists a set $E \varepsilon \mathfrak{B}$ which is disjoint with a set $E_1 \varepsilon \mathfrak{B}_1$. Since E and E_1 are closed and X is normal there is a bounded continuous real-valued function f defined on X which has the value one for $t \varepsilon E$, zero for $t \varepsilon E_1$. Defining $g(\cdot) \varepsilon C(X, A)$ by $g(t) = f(t)e$ (e being the unit of A) we have from (4.3) using $g(\cdot)$ that $1 = 0$ which is impossible. Hence $\mathfrak{B} = \mathfrak{B}_1$.

On the basis of the argument of Lemma 4. 8 we may say there is just one $0 - 1$ measure function μ associated with the maximal ideal N of $C(X, A)$ such that N is the set of zeros of a multiplicative linear functional of the form (4. 2).

4. 9. LEMMA. *Let N be a maximal ideal of $C(X, A)$ and let μ be the associated $0 - 1$ measure function. Then the collection $\Sigma(N)$ of all $\gamma \varepsilon \Gamma$ for which N is the set of zeros of $\int_X f(t, \gamma(t)) d\mu$ in $C(X, A)$ is non-empty and is $\bigcap \Gamma(f, \epsilon)$, $f \varepsilon N$.*

That the set $\Sigma(N)$ in question is non-empty follows from Lemma 4. 7. Suppose $\gamma \varepsilon \bigcap \Gamma(f, \epsilon)$, $f \varepsilon N$, which we note is independent of the choice of $\epsilon > 0$. Then we have

$$\left| \int_X f(t, \gamma(t)) d\mu \right| \leqq \int_X | f(t, \gamma(t)) | \, d\mu \leqq \epsilon$$

for each $\epsilon > 0$, $f \varepsilon N$, and hence $\gamma \varepsilon \Sigma(N)$. Conversely if $\gamma \varepsilon \Sigma(N)$, then for each $f(\cdot) \varepsilon N$ there is by the arguments used for Lemma 4. 6 (see [18, Lemma 2. 4]) a function $g(\cdot) \varepsilon N$ such that $g(t, M) = | f(t, M) |$ for each $t \varepsilon X$, $M \varepsilon \mathfrak{M}$. Since

$$\int_X g(t, \gamma(t)) d\mu = \int_X | f(t, \gamma(t)) | \, d\mu = 0$$

it follows that for each $\epsilon > 0$, $\gamma \epsilon \Gamma(f, \epsilon)$. Since $f(\cdot)$ is arbitrary in N, $\gamma \varepsilon \bigcap \Gamma(f, \epsilon)$.

4. 10. THEOREM. *To each maximal additive ideal $\mathfrak{B}$ in $\mathfrak{L}$ there corresponds a decomposition of Γ into disjoint sets $\Gamma(\mathfrak{B})$ such that the maximal ideals of $C(X, A)$ are in a 1-1 correspondence with the pairs $(\mathfrak{B}, \delta)$, δ being in $\Gamma(\mathfrak{B})$. This correspondence has the property that if δ' is any element of δ and if μ is the 0-1 measure function associated with $\mathfrak{B}$, then the maximal ideal which corresponds to $(\mathfrak{B}, \delta)$ is the set of $f(\cdot)$ for which*

$$\int_X f(t, \delta'(t)) d\mu = 0. \tag{4. 4}$$

Let $\mathfrak{B}$ be a m. a. ideal of $\mathfrak{L}$ and μ be its associated 0-1 measure function and let γ and γ_1 be in Γ. We say that $\gamma \sim \gamma_1$ (with respect to $\mathfrak{B}$) if the relation (4. 3) holds (with $\mu_1 = \mu$). Clearly $\sim$ is an equivalence relation defined on Γ and provides a decomposition $\Gamma(\mathfrak{B})$ of Γ. If δ is such an equivalence class and γ and γ_1 are in δ, then the zeros of the integrals of (4. 3) constitute the same maximal ideal N. That a maximal ideal is so obtained follows from the fact that the integral is a multiplicative linear functional on $C(X, A)$. Then by Lemma 4. 9, $\delta = \bigcap \Gamma(f, \epsilon)$, $f \varepsilon N$.

Given the maximal ideal N of $C(X, A)$ we correspond to it the pair $(\mathfrak{B}, \delta)$ where $\mathfrak{B}$ is the sole m. a. ideal of $\mathfrak{L}$ which contains $\mathfrak{A}(N)$ by Lemma 4.3 and $\delta = \bigcap \Gamma(f, \epsilon)$, $f \varepsilon N$. N is the set of zeros of (4.4) by Lemma 4.9. No two distinct maximal ideals can correspond to the same pair $(\mathfrak{B}, \delta)$. Also if $(\mathfrak{B}, \delta)$ is given it is clear from the above that there is precisely one maximal ideal N of $C(X, A)$ which maps into $(\mathfrak{B}, \delta)$ under the correspondence.

We seek a criterion not involving a knowledge of the maximal ideals of $C(X, A)$ to determine whether or not $\gamma \sim \gamma_1$ (with respect to $\mathfrak{B}$). This we are able to do under the additional assumption that $X \times \mathfrak{M}$ is normal.

4.11. LEMMA. *Let $X \times \mathfrak{M}$ be normal (as well as X). Then $\gamma_1 \sim \gamma_2$ (with respect to $\mathfrak{B}$) if and only if the closures of the sets $\{(t, \gamma_i(t)) \mid t \varepsilon F\}$, $i = 1, 2$ in $X \times \mathfrak{M}$ are not disjoint for each $F \varepsilon \mathfrak{B}$.*

Suppose that $\gamma_1 \not\sim \gamma_2$. Then there exists $f(\cdot) \varepsilon C(X, A)$ for which

$$\int_X f(t, \gamma_i(t))\, d\mu = \begin{cases} 0, & i = 1 \\ 1, & i = 2 \end{cases}. \tag{4.5}$$

Let $f'(\cdot)$ be the adjoint of $f(\cdot)$, $f'(t, M) \equiv \overline{f(t, M)}$. Then by the facts that the integral is a multiplicative linear functional and that the integrals for $f'(\cdot)$ have values conjugate to those for $f(\cdot)$, it follows that the relations (4.5) hold for $g(\cdot) = f(\cdot) f'(\cdot)$ in place of $f(\cdot)$. Note that for each (t, M), $g(t, M) \geqq 0$. Then by the continuity of the integrand as a function of t, its positivity and the fact that the integral is zero we see that

$F_1 = \{t \varepsilon X \mid |g(t, \gamma_1(t))| \leqq 1/4\}$ is a closed set and $\mu(F_1) = 1$.

By the same reasoning, the set $F_2 = \{t \varepsilon X \mid |g(t, \gamma_2(t))| \geqq 1/2\}$ is a closed set, $\mu(F_2) = 1$.

Let $F = F_1 \cap F_2$. Since each $\mu(F_i) = 1$ and μ is a 0-1 measure function on $\mathfrak{S}$, by Lemma 4.4, $\mu(F) = 1$ and F intersects every set in $\mathfrak{B}$. Thus $F \varepsilon \mathfrak{B}$, and is non-empty.

Consider the sets $S_i = \{(t, \gamma_i(t)) \mid t \varepsilon F_1 \cap F_2\}$, $i = 1, 2$. If (t_0, M_0) is in the intersection of their closures in $X \times \mathfrak{M}$ it is readily verified by Lemma 2.2 that $|g(t_0, M_0)| \leqq 1/4$, $|g(t_0, M_0)| \geqq 1/2$ which is a contradiction.

For the converse we suppose that there exists a set $F \varepsilon \mathfrak{B}$, such that the closures in $X \times \mathfrak{M}$ of the sets $S_i = \{(t, \gamma_i(t)) \mid t \varepsilon F\}$, $i = 1, 2$ are disjoint. Since $X \times \mathfrak{M}$ is normal, there is a bounded positive continuous function on $X \times \mathfrak{M}$ which has the value zero for S_1, one for S_2. Let $f(\cdot)$ be the corresponding member of $C(X, A)$ by Theorem 3.1. Then the relations (4.5) hold since $\mu(F) = 1$. This shows that $\gamma_1 \not\sim \gamma_2$.

Thus we have liberated the equivalence relation $\sim$ (with respect to $\mathfrak{B}$) from the maximal ideals of $C(X, A)$ and have it described entirely in terms of X and $\mathfrak{M}$. Next we start with the normal space X, and the compact space $\mathfrak{M}$ where $X \times \mathfrak{M}$ is normal and consider the B^*-algebra $C(X \times \mathfrak{M})$ of complex bounded continuous functions on $X \times \mathfrak{M}$. By Theorem 3.1 this is equivalent to $C(X, C(\mathfrak{M}))$. The equivalence classes of $\sim$ are families of functions defined on X with values in $\mathfrak{M}$. To obtain the β-compactification of $X \times \mathfrak{M}$ we describe the topology of the space of maximal ideals of the latter algebra.

But first we note that if $(\mathfrak{B}, \delta)$ is a pair in Theorem 4.10 with the associated set function μ, then if $f(\cdot)$ satisfies (4.4) so does its adjoint $f'(\cdot)$. Since the integral is a multiplicative functional and preserves complex conjugation, (4.4) holds for $f(\cdot)$ if and only if it holds for $f(\cdot)f'(\cdot)$. Thus $\mathfrak{A}(f, \gamma, \epsilon) \varepsilon \mathfrak{B}$ for each $\epsilon > 0$, $\gamma \varepsilon \delta$ is equivalent to (4.4) (see the discussion preceding Lemma 4.7 for the definition of $\mathfrak{A}(f, \gamma, \epsilon)$).

4.12. THEOREM. *Let X and $X \times \mathfrak{M}$ be normal where $\mathfrak{M}$ is compact. The points of the β-compactification of $X \times \mathfrak{M}$ are in a 1-1 correspondence with the pairs $(\mathfrak{B}, \delta)$ where $\mathfrak{B}$ is a m. a. ideal of $\mathfrak{L}$ and δ is an equivalence class of $\sim$ with respect to $\mathfrak{B}$. Let E be any set in the β-compactification. Then its closure $\bar{E}$ there is the set of all $(\mathfrak{B}_1, \delta_1)$ for which $\mathfrak{A}(f, \gamma, \epsilon) \varepsilon \mathfrak{B}_1$ for each $\epsilon > 0$, $\gamma_1 \varepsilon \delta_1$ whenever $\mathfrak{A}(f, \gamma, \epsilon) \varepsilon \mathfrak{B}$ for each $\epsilon > 0$, $(\mathfrak{B}, \delta) \varepsilon E, \gamma \varepsilon \delta$.*

Since the maximal ideals are given as the zeros of corresponding Hildebrandt integrals (Theorem 4.10), by the previous discussion this is a description of closure in the space of maximal ideals of $C(X, C(\mathfrak{M}))$ using the Stone topology (see § 2). Thus we have succeeded in describing the β-compactification in terms of the β-compactification of X, $P_{t \varepsilon X}\mathfrak{M}$, continuous functions $X \times \mathfrak{M}$ and closure in $X \times \mathfrak{M}$. That the m. a. ideals of $\mathfrak{L}$ constitute the β-compactification of X is assured since X is normal.

Concerning the relation of $(\mathfrak{B}_1, \delta_1) \varepsilon \bar{E}$ to E we can say a bit more. Let $\mathfrak{E}$ be the additive ideal of the lattice $\mathfrak{L}$ which consists of all closed subsets of $\mathfrak{L}$ common to all the $\mathfrak{B}$ of $(\mathfrak{B}, \delta)$ in E. $\mathfrak{E}$ is non-empty as it contains at least the set X.

4.14. THEOREM. *In the notation of Theorem 4.13, if $(\mathfrak{B}_1, \delta_1) \varepsilon \bar{E}$, then $\mathfrak{E}$ is contained in $\mathfrak{B}_1$.*

Suppose that there exists a set F in $\mathfrak{E}$ which is not in $\mathfrak{B}_1$. Since $\mathfrak{B}_1$ is a m. a. ideal, there is a set $F_1 \varepsilon \mathfrak{B}_1$ which is disjoint with F. By the normality of X, there is a continuous bounded real-valued function $f(t)$ which is one on F_1 and zero on F. Then let $g(t) = f(t)e$. We have

$$\int_X g(t,\gamma(t))\,d\mu = \int_F g(t,\gamma(t))\,d\mu = 0$$

for each μ associated with $\mathfrak{B}$ of $(\mathfrak{B},\delta) \varepsilon E$, $\gamma \varepsilon \delta$. Also the integral for μ_1 associated with B_1 and $\gamma_1 \varepsilon \delta_1$ has the value 1. This is impossible as $(B_1, \delta_1) \varepsilon \bar{E}$.

5. *X* compact, *A* a Banach algebra. In this section we are concerned with the space of maximal ideals of $C(X,A)$ when X is a compact Hausdorff space and A is a Banach algebra but not a B^*-algebra. In particular it is shown that the space of maximal ideals is homeomorphic to X if A is a primary ring in the sense of Šilov [13], that is, possesses one maximal ideal.

5.1. LEMMA. *Let X be a compact Hausdorff space and every maximal ideal in $C(X,A)$ be fixed. Then the space of maximal ideals $\mathfrak{N}$ of $C(X,A)$ is homeomorphic to $X \times \mathfrak{M}$.*

It is easy to verify that if $t_i \varepsilon X$, $M_i \varepsilon \mathfrak{M}$, $i = 1, 2$, then the maximal ideals $\{f(\cdot) \mid f(t_i) \varepsilon M_i\}$, $i = 1, 2$ are the same if and only if $t_1 = t_2$ and $M_1 = M_2$. Thus $\mathfrak{N}$ may be considered as the set of all pairs (t, M), $t \varepsilon X$, $M \varepsilon \mathfrak{M}$ supplied with an appropriate topology. But by Lemma 2.2, for each $f(\cdot) \varepsilon C(X,A)$, $f(t,M)$ is continuous on $X \times \mathfrak{M}$ which is a compact Hausdorff space. By a theorem of Gelfand [3], $\mathfrak{N}$ is homeomorphic to $X \times \mathfrak{M}$.

5.2. THEOREM. *Let X be a compact Hausdorff space and A be a primary ring. Then the space $\mathfrak{N}$ of maximal ideals of $C(X,A)$ is homeomorphic to X.*

By Lemma 5.1 this result holds if every maximal ideal $N \varepsilon \mathfrak{N}$ is fixed. Suppose that N is not fixed. Let M be the maximal ideal of A. Then for each $t_0 \varepsilon X$ there exists $f(\cdot) \varepsilon N$ such that $f(t_0, M) \neq 0$. Since $f(t, M)$ is continuous in t by Lemma 2.2, there exists a neighborhood $U(t_0)$ of t_0 in X where $f(t,M) \neq 0$. A finite number of these neighborhoods suffice to cover X. Let $f_1(\cdot), \cdots, f_n(\cdot)$ be the corresponding elements of N. We define $g_i(\cdot)$ by $g_i(t) = \overline{f_i(t,M)}e$ for each $t \varepsilon X$, $i = 1, \cdots, n$. Then $g_i(\cdot) \varepsilon C(X,A)$, $i = 1, \cdots, n$. Set $h(\cdot) = \sum_{i=1}^{n} f_i(\cdot) g_i(\cdot)$. It is easy to see that for each $t \varepsilon X$, $h(t,M) \neq 0$ and that $h(\cdot) \varepsilon N$. For each $t \varepsilon X$, $[h(t)]^{-1}$ exists in A ([3]). But $[h(t)]^{-1}$ as a function of t is a continuous function of a continuous function defined on X ([9, Theorem 5.2.3]) and moreover is bounded since X is compact. Thus $h(\cdot)$ has an inverse in $C(X,A)$ which contradicts the fact that $h(\cdot)$ is in the maximal ideal N.

CORNELL UNIVERSITY.

REFERENCES.

[1] Garrett Birkhoff, *Lattice theory*, American Mathematical Society Colloquium Publications, vol. 25, New York, 1940.

[2] Eduard Čech, "On bicompact spaces," *Annals of Mathematics*, vol. 38 (1937), pp. 823-844.

[3] I. Gelfand, "Normierte Ringe," *Recueil Mathématique* (*Mat. Sbornik*), N. S., vol. 9 (1941), pp. 3-24.

[4] I. Gelfand and A. N. Kolmogoroff, "On rings of continuous functions on topological spaces," *Comptes Rendus* (*Doklady*) *de l'Académie des Sciences de l'URSS*, vol. 22 (1939), pp. 11-15.

[5] I. Gelfand and M. Neumark, "On the imbedding of normed rings into the ring of operators in Hilbert space," *Recueil Mathématique* (*Mat. Sbornik*), N. S., vol. 12 (1943), pp. 197-213.

[6] I. Gelfand and G. Šilov, "Über verschiedene Methoden der Einführung der Topologie in die Menge der maximalen Idealen eines normierten Ringes," *Recueil Mathématique* (*Mat. Sbornik*), N. S., vol. 9 (1941), pp. 25-28.

[7] Edwin Hewitt, "Rings of real-valued continuous functions. I," *Transactions of the American Mathematical Society*, vol. 64 (1948), pp. 45-99.

[8] T. H. Hildebrandt, "On bounded linear functional operations," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 419-427.

[9] Einar Hille, *Functional Analysis and Semi-groups*, American Mathematical Society Colloquium Publications, vol. 31, New York, 1948.

[10] Irving Kaplansky, "Topological rings," *American Journal of Mathematics*, vol. 69 (1947), pp. 153-183.

[11] C. E. Rickart, "Banach algebras with an adjoint operation," *Annals of Mathematics*, vol. 47 (1946), pp. 528-550.

[12] G. Šilov, "On the extension of maximal ideals," *Comptes Rendus* (*Doklady*) *de l'Académie des Sciences de l'URSS*, vol. 29 (1940), pp. 83-84.

[13] G. Šilov, "On regular normed rings," *Travaux de l'Institut Mathématique Slekloff*, vol. 21 (1947). (Russian. English summary.)

[14] V. Šmulian, "On multiplicative linear functionals in certain special normed rings," *Comptes Rendus* (*Doklady*) *de l'Académie des Sciences de l'URSS*, vol. 26 (1940), pp. 13-16.

[15] M. H. Stone, "Applications of the theory of Boolean rings to general topology," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 375-481.

[16] M. H. Stone, "Topological representations of distributive lattices and Brouwerian logics," *Časopis pro Pěstování Matematiky a Fysiky*, vol. 67 (1937), pp. 1-25.

[17] Henry Wallman, "Lattices and topological spaces," *Annals of Mathematics*, vol. 39 (1938), pp. 112-126.

[18] Bertram Yood, "Banach algebras of bounded functions," *Duke Mathematical Journal*, vol. 16 (1949), pp. 151-163.

A FORM OF GREEN'S TRANSFORMATION.*

By C. Truesdell.

The transformation of a volume integral into a surface integral familiar in the equivalent forms

$$(1)\qquad \oint_S d\boldsymbol{S}\boldsymbol{b} = \oint_V \mathbf{grad}\,\boldsymbol{b}\,d\mathfrak{V}, \quad \oint_S d\boldsymbol{S}\cdot\boldsymbol{b} = \oint_V \operatorname{div}\boldsymbol{b}\,d\mathfrak{V},$$

$$\oint_S d\boldsymbol{S}\times\boldsymbol{b} = \oint_V \mathbf{curl}\,\boldsymbol{b}\,d\mathfrak{V},$$

which is known generally as "Gauss's theorem," "Ostrogradski's theorem," "Green's theorem," or the "divergence theorem," was first employed in a special case by Lagrange,[1] and its systematic use in mathematical physics was initiated by Green,[2] who discovered the equivalent forms

$$(2)\qquad \begin{aligned} \oint_S d\boldsymbol{S}\cdot(\phi\,\mathbf{grad}\,\psi - \psi\,\mathbf{grad}\,\phi) &= \oint_V (\phi\nabla^2\psi - \psi\nabla^2\phi)\,d\mathfrak{V}, \\ \oint_S d\boldsymbol{S}\cdot\phi\,\mathbf{grad}\,\psi &= -\oint_V (\phi\nabla^2\psi + \mathbf{grad}\,\phi\cdot\mathbf{grad}\,\psi)\,d\mathfrak{V}. \end{aligned}$$

There are many other forms of this transformation. Any form containing at least as many arbitrary functions of position as the order of the highest derivative occurring in it is equivalent to all the others, and no new results can be obtained by building up apparently more complicated combinations. Some of these, however, are useful in mechanics and physics, so that various books and papers pursue the manipulations necessary to obtain them. The purpose of the present note is to give an economical derivation of a new form sufficiently elaborate to include and relate a number of known examples.

For the polyadic n-th power[3] of a vector $\boldsymbol{b}$ let us introduce the notation $\boldsymbol{b}^{(n)}$:

---

* Received October 12, 1949.

[1] J. L. Lagrange, *Œuvres*, vol. 1, pp. 151-316; see § 45. In the work of Gauss also only special cases appear: *Gesammelte Werke*, vol. 5, pp. 3-22, see§§ 3-5, and pp. 195-242, see § 22.

[2] G. Green, *Mathematical Papers*, pp. 23-26.

[3] Our polyadic notation is essentially that of Gibbs and Wilson, *Vector Analysis*, New Haven (1902). In indicial notation $b^{(n)}$ is $b^{i_1}b^{i_2}\cdots b^{i_n}$.

$$\boldsymbol{b}^{(-1)} \equiv 0, \quad \boldsymbol{b}^{(0)} \equiv 1, \quad \boldsymbol{b}^{(1)} \equiv \boldsymbol{b}, \quad \boldsymbol{b}^{(2)} \equiv \boldsymbol{bb}, \quad \boldsymbol{b}^{(3)} \equiv \boldsymbol{bbb}, \cdots. \tag{3}$$

Let $\boldsymbol{\Sigma}$ be any polyadic (i. e., tensor), and let the notation $\{\boldsymbol{\Sigma b}^{(n)}\}$ stand for the polyadic[4]

$$\{\boldsymbol{\Sigma b}^{(n)}\} \equiv \boldsymbol{b}^{(n)}\boldsymbol{\Sigma} + \boldsymbol{b}^{(n-1)}\boldsymbol{\Sigma b} + \boldsymbol{b}^{(n-2)}\boldsymbol{\Sigma b}^{(2)} + \cdots + \boldsymbol{\Sigma b}^{(n)}, \tag{4}$$

which is completely symmetrical with respect to $\boldsymbol{b}$.

Let $\boldsymbol{b}$ and $\boldsymbol{c}$ be arbitrary vector fields, and let $\boldsymbol{r}$ be the radius vector. First, since $\mathbf{grad}\, \boldsymbol{r} = \boldsymbol{I}$, we have

$$\operatorname{div}(\boldsymbol{bcr}^{(n)}) = \boldsymbol{c}\{\boldsymbol{br}^{(n-1)}\} + (\boldsymbol{b}\cdot\mathbf{grad}\, \boldsymbol{c})\boldsymbol{r}^{(n)} + (\boldsymbol{c} \operatorname{div} \boldsymbol{b})\boldsymbol{r}^{(n)}. \tag{5}$$

Similarly

$$\mathbf{grad}(\boldsymbol{b}\cdot\boldsymbol{cr}^{(n)}) = \boldsymbol{b}\cdot\boldsymbol{c}\{\boldsymbol{r}^{(n-1)}\boldsymbol{I}\} + (\mathbf{grad}\, \boldsymbol{c}\cdot\boldsymbol{b} + \mathbf{grad}\, \boldsymbol{b}\cdot\boldsymbol{c})\boldsymbol{r}^{(n)}. \tag{6}$$

Since

$$\boldsymbol{b}\cdot\mathbf{grad}\, \boldsymbol{c} - \mathbf{grad}\, \boldsymbol{c}\cdot\boldsymbol{b} = \boldsymbol{curl}\, \boldsymbol{c} \times \boldsymbol{b}, \tag{7}$$

from (5) and (6) we obtain

$$\begin{aligned} &\operatorname{div}[(\boldsymbol{bc} + \boldsymbol{cb})\boldsymbol{r}^{(n)}] - \mathbf{grad}(\boldsymbol{b}\cdot\boldsymbol{cr}^{(n)}) \\ &\quad = \boldsymbol{c}\{\boldsymbol{br}^{(n-1)}\} + \boldsymbol{b}\{\boldsymbol{cr}^{(n-1)}\} - \boldsymbol{b}\cdot\boldsymbol{c}\{\boldsymbol{r}^{(n-1)}\boldsymbol{I}\} \\ &\qquad + (\mathbf{curl}\, \boldsymbol{c} \times \boldsymbol{b} + \mathbf{curl}\, \boldsymbol{b} \times \boldsymbol{c} + \boldsymbol{b} \operatorname{div} \boldsymbol{c} + \boldsymbol{c} \operatorname{div} \boldsymbol{b})\boldsymbol{r}^{(n)}. \end{aligned} \tag{8}$$

Now the familiar forms (1) remain valid when $\boldsymbol{b}$ is replaced by a polyadic (tensor) $\boldsymbol{\Sigma}$ of any rank. Hence integrating (8) over a volume yields

$$\begin{aligned} &\oint_S [d\boldsymbol{S}\cdot(\boldsymbol{bc} + \boldsymbol{cb})\boldsymbol{r}^{(n)} - d\boldsymbol{Sb}\cdot\boldsymbol{cr}^{(n)}] \\ &\quad = \int_V [\boldsymbol{b}\{\boldsymbol{cr}^{(n-1)}\} + \boldsymbol{c}\{\boldsymbol{br}^{(n-1)}\} - \boldsymbol{b}\cdot\boldsymbol{c}\{\boldsymbol{r}^{(n-1)}\boldsymbol{I}\} \\ &\qquad + (\mathbf{curl}\, \boldsymbol{c} \times \boldsymbol{b} + \mathbf{curl}\, \boldsymbol{b} \times \boldsymbol{c} + \boldsymbol{b} \operatorname{div} \boldsymbol{c} + \boldsymbol{c} \operatorname{div} \boldsymbol{b})\boldsymbol{r}^{(n)}]\, d\mathfrak{V}, \end{aligned} \tag{9}$$

which is the desired result.

By putting $n = 0$ we obtain the transformation

$$\begin{aligned} &\oint_S [d\boldsymbol{S}\cdot(\boldsymbol{bc} + \boldsymbol{cb}) - d\boldsymbol{S}\, \boldsymbol{b}\cdot\boldsymbol{c}] \\ &\quad = \int_V [\mathbf{curl}\, \boldsymbol{c} \times \boldsymbol{b} + \mathbf{curl}\, \boldsymbol{b} \times \boldsymbol{c} + \boldsymbol{b} \operatorname{div} \boldsymbol{c} + \boldsymbol{c} \operatorname{div} \boldsymbol{b}]\, d\mathfrak{V}, \end{aligned} \tag{10}$$

---

[4] In indicial notation $\{\boldsymbol{\Sigma}\, \boldsymbol{b}^{(n)}\}$ is $\sum_{k=0}^{n} [(\prod_{j=1}^{k} b^{i_j}) \Sigma^{lm\ldots p} (\prod_{j=k+1}^{n} b^{i_j})]$.

discovered by Burgatti,[5] who showed it to contain as special cases a formula of Levi-Civita in the theory of liquid jets, a formula of Teofilato in aerodynamics and a formula of Vaschy in electromagnetic theory, as well as various classical formulae of potential theory. Earlier, however, Lamb [6] and J. J. Thompson [7] in their studies of the rotational motions of fluids had derived the special case when $\boldsymbol{c} = \boldsymbol{b} =$ the velocity of a motion of an incompressible substance ($\operatorname{div} \boldsymbol{b} = 0$); Lamb's result is that if $\lim r^2 b = 0$ as $r \to \infty$ then the integral $\int_V \mathbf{curl}\, \boldsymbol{b} \times \boldsymbol{b} d\mathfrak{V}$ taken over the whole of a continuous motion filling all space must vanish.

By putting $n = 1$ in our general identity (9) we obtain

$$\begin{aligned} &\oint_S [d\boldsymbol{S} \cdot (\boldsymbol{bc} + \boldsymbol{cb})\boldsymbol{r} - d\boldsymbol{S}\boldsymbol{b} \cdot \boldsymbol{cr}] \\ (11) \qquad &= \int_V [\boldsymbol{bc} + \boldsymbol{cb} - \boldsymbol{b} \cdot \boldsymbol{c}\boldsymbol{I} + (\mathbf{curl}\, \boldsymbol{c} \times \boldsymbol{b} + \mathbf{curl}\, \boldsymbol{b} \times \boldsymbol{c} + \boldsymbol{b} \operatorname{div} \boldsymbol{c} + \boldsymbol{c} \operatorname{div} \boldsymbol{b})\boldsymbol{r} d\mathfrak{V}. \end{aligned}$$

Taking the scalar of this dyadic equation [8] yields

$$\begin{aligned} &\oint_S [d\boldsymbol{S} \cdot (\boldsymbol{bc} + \boldsymbol{cb}) \cdot \boldsymbol{r} - d\boldsymbol{S} \cdot \boldsymbol{r}\boldsymbol{b} \cdot \boldsymbol{c}] \\ (12) \qquad &= \int_V [-\boldsymbol{b} \cdot \boldsymbol{c} + \boldsymbol{r} \cdot (\mathbf{curl}\, \boldsymbol{c} \times \boldsymbol{b} + \mathbf{curl}\, \boldsymbol{b} \times \boldsymbol{c} + \boldsymbol{b} \operatorname{div} \boldsymbol{c} + \boldsymbol{c} \operatorname{div} \boldsymbol{b})] d\mathfrak{V}. \end{aligned}$$

Put $\boldsymbol{c} = \boldsymbol{b}$ and suppose $\operatorname{div} \boldsymbol{b} = 0$; then

$$(13) \qquad \tfrac{1}{2} \int_V b^2 d\mathfrak{V} = \oint_S d\boldsymbol{S} \cdot (\tfrac{1}{2}\boldsymbol{I}b^2 - \boldsymbol{bb}) \cdot \boldsymbol{r} + \int_V \boldsymbol{r} \cdot \mathbf{curl}\, \boldsymbol{b} \times \boldsymbol{b}\, d\mathfrak{V}.$$

When $\boldsymbol{b}$ is the velocity vector of a motion, $\mathbf{curl}\, \boldsymbol{b}$ is the vorticity; for an incompressible substance, since $\operatorname{div} \boldsymbol{b} = 0$ we may apply (13), and if further the susbtance be of uniform density, the left side is proportional to the total kinetic energy, and we obtain another result of Lamb [9] and J. J. Thompson.[10]

---

[5] P. Burgatti, "Intorno a una formula generale di trasformazione di un integrale di spazio in uno di superficie e alle sue varie deduzioni," *Bolletino della Unione Matematica Italiana*, vol. 10 (1931), pp. 1-5.

[6] H. Lamb, *A Treatise on the Mathematical Theory of the Motion of Fluids*, Cambridge (1879), see § 136. Also *Hydrodynamics*, 6th ed., Cambridge (1932), § 153.

[7] J. J. Thomson, *A Treatise on the Motion of Vortex Rings*, London (1883). See § 4.

[8] In tensorial language, contracting the equation $\Phi^{i}{}_{j} = \dots$; operationally, inserting a dot into each dyad of the dyadic.

[9] *Loc. cit.*

[10] *Op. cit.*, § 6.

From this identity we notice also that if $\mathbf{curl}\, \boldsymbol{b} \times \boldsymbol{b} = 0$ everywhere and if $\boldsymbol{b} = 0$ upon the boundary surface $\mathcal{S}$, then both integrals on the right side vanish and hence $\boldsymbol{b} = 0$. Thus we have an important theorem stated in principle but not adequately proved by Duhem:[11] There can exist no continuously differentiable motion of an incompressible substance such that (i) the vorticity is everywhere zero or parallel to the velocity, (ii) the material adheres to all finite boundaries without slipping, and in infinite regions $\lim r^3 b^2 = 0$ as $r \to \infty$. For a second application, put $\boldsymbol{b} = \mathbf{curl}\, \boldsymbol{v}$, where $\boldsymbol{v}$ is the velocity vector of a motion of any substance, so that $\operatorname{div} \boldsymbol{b} = 0$ again. From (13) we deduce the following result: In any motion such that the vorticity is everywhere parallel to its own curl, if the vorticity $\boldsymbol{b}$ vanish upon all finite boundaries and $\lim r^3 b^2 = 0$ as $r \to \infty$ in all infinite regions, the vorticity itself must vanish everywhere. This theorem is interesting because the viscous forces in a liquid of uniform viscosity depend upon $\mathbf{curl}\, \boldsymbol{b}$ only, so that when $\mathbf{curl}\, \boldsymbol{b} = 0$ even if $\boldsymbol{b} \neq 0$, the equations of motion reduce to those of perfect fluids; our result shows that such simple motions cannot be imbedded continuously in a region whose boundaries are in irrotational motion. Returning to the dyadic identity (11), we may obtain another form by taking its vector or cross[12], viz.,

$$(14) \quad \oint_{\mathcal{S}} [d\boldsymbol{S} \cdot (\boldsymbol{bc} + \boldsymbol{cb}) \times \boldsymbol{r} - d\boldsymbol{S} \times \boldsymbol{rb} \cdot \boldsymbol{b}] = \int_{\mathcal{V}} (\mathbf{curl}\, \boldsymbol{c} \times \boldsymbol{b} + \mathbf{curl}\, \boldsymbol{b} \times \boldsymbol{c} + \boldsymbol{b} \operatorname{div} \boldsymbol{c} + \boldsymbol{c} \operatorname{div} \boldsymbol{b}) \times \boldsymbol{r} \, d\mathfrak{V}.$$

Again put $\boldsymbol{c} = \boldsymbol{b}$ and suppose $\operatorname{div} \boldsymbol{b} = 0$; the resulting special case is

$$(15) \quad \oint_{\mathcal{S}} [d\boldsymbol{S} \cdot \boldsymbol{br} \times \boldsymbol{b} + \tfrac{1}{2} b^2 d\boldsymbol{S} \times \boldsymbol{r}] = \int_{\mathcal{V}} \boldsymbol{r} \times (\mathbf{curl}\, \boldsymbol{b} \times \boldsymbol{b}) \, d\mathfrak{V},$$

an identity used by Poincaré[13] in his analysis of the vorticity of fluid motions.

---

[11] P. Duhem, "Recherches sur l'hydrodynamique, 5e partie, Le théorème de Lagrange et les conditions aux limites," *Annales de la Faculté des Sciences de Toulouse*, ser. 2, vol. 5 (1903), pp. 353-376; see Chapter II, § 1. Duhem considered only the case $\mathbf{curl}\, \boldsymbol{b} = 0$. The more general result for "Beltrami motions" ($\mathbf{curl}\, \boldsymbol{b} \times \boldsymbol{b} = 0$) is here stated and proved for the first time. Duhem's theorem is not to be confused with a somewhat similar result in potential theory, where the condition $\boldsymbol{b} = 0$ upon $\mathcal{S}$ is not required, but the domain must be simply connected.

[12] "Taking the cross" of $\Phi^{i}{}_{j} = \ldots$ means forming $\epsilon_{ijk}\Phi^{jk} = \ldots$; operationally, inserting a cross into each dyad of the dyadic.

[13] H. Poincaré, *Théorie des tourbillons*, Paris (1893), § 115.

The occurrence of $r^{(n)}$ in the general transformation (9) suggests that the vectorial moments of a vector field, particularly of a solenoidal vector field, have simple and useful properties. Particular cases, such as polyadic cross moments $\boldsymbol{r} \times (\boldsymbol{r} \times (\boldsymbol{r} \times \cdots (\boldsymbol{r} \times \boldsymbol{\Sigma}) \cdots))$ and special types of scalar moments like $\boldsymbol{r} \cdot (\boldsymbol{r} \times (\boldsymbol{r} \cdot (\boldsymbol{r} \times \boldsymbol{\Sigma})))$ for a dyadic $\boldsymbol{\Sigma}$, suggest themselves at once. In a later paper[14] I shall show that the study of vectorial moments enables the method introduced by Poincaré in the theory of vortices to be generalized so as to apply to any deformable continuum in motion.

APPLIED MATHEMATICS BRANCH, MECHANICS DIVISION,
NAVAL RESEARCH LABORATORY
AND
DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MARYLAND.

---

[14] C. Truesdell, "Vorticity averages," *Canadian Journal of Mathematics* (in press).

DIFFERENTIAL EQUATIONS REFERRED TO A VARIABLE METRIC.*

By D. C. Lewis.

1. Introduction. In a previous paper,[1] hereafter denoted by MP, the author studied the dependence of solutions of systems of differential equations on the initial conditions with the help of a general Finsler metric, which, of course, includes as a special case a Riemann metric. In applications to exterior ballistics and other problems, the success of this method is naturally seen to depend on the choice of the metric, but it was disappointing that for *no* choice of the metric was it possible to deduce from this theory certain well known results about even linear differential equations. For example, it appeared to be impossible to deduce by this method that any solution of the Bessel differential equation is $O(t^{-\frac{1}{2}})$, where t is the independent variable. It thus was felt that the method as so far developed was intrinsically too crude to yield the desired results in applications.

The present paper makes at least a first step in remedying this situation. This is done by introducing a metric which varies not only from point to point of the space, but also varies with the independent variable t of the differential system. The resulting theorems seem to be but trivial modifications of those in MP, and yet the power of the method when applied to linear differential equations is greatly increased, as is shown in Section **3**, where some results of Wintner[2] and Leighton[3] are obtained by this method together with the well known result on Bessel functions previously alluded to.

The success of the method still depends on the choice of the metric, and a general theory of such a choice is still unavailable. Nevertheless, the possibility of approximate linearization in many notable cases indicates the essential importance of the fairly complete analysis of this question in the linear case, appearing in Section **5**.

---

* Received June 1, 1950.

[1] D. C. Lewis, "Metric properties of differential equations," *American Journal of Mathematics*, vol. 71 (1949), pp. 294-312.

[2] A. Wintner, "Stability and high frequency," *Journal of Applied Physics*, vol. 18 (1947), pp. 941-942.

[3] W. Leighton, "Bound for the solutions of a second-order linear differential equation," *Proceedings of the National Academy of Sciences*, vol. 35 (1949), pp. 190-191.

2. The fundamental theorems referred to a Finsler variable metric. The idea of a variable metric is simply that the distance between two points is to be specified only when a variable t, which we may think of as representing time, takes on a definite value. If we are thinking of a Riemann metric $\sum_{ij} g_{ij}dx_i dx_j$, this means that the g's are not only functions of $x_1, \cdots, x_n$ but also are functions of t. It means that the "distance" between two points changes in general with t, even if the points themselves are fixed. In the immediate sequel we consider the somewhat more general case of a Finsler metric.

Let $f[x, \dot{x}, t] = f(x_1, \cdots, x_n, \dot{x}_1, \cdots, \dot{x}_n, t)$ be of class C' in all of its $2n + 1$ arguments and suppose that, for each fixed value of t, it is of class C''' in the x's and $\dot{x}$'s. Furthermore we assume $f[x, \dot{x}, t]$ to be positively homogeneous of degree $+1$ in $\dot{x}_1, \cdots, \dot{x}_n$, and that it is positive unless all the $\dot{x}$'s are zero. Finally, in order to insure the "regularity" of the variational problem $\delta \int f[x, \dot{x}, t]d\tau = 0$, we assume that $f_{\dot{x}_i\dot{x}_j}[x, \dot{x}, t]\lambda^i\lambda^j > 0$ for all sets $\lambda^1, \cdots, \lambda^n$ not proportional to $\dot{x}_1, \cdots, \dot{x}_n$. Here, as in the sequel, the summation convention of tensor analysis is used, with all repeated indices summed from 1 to n.

The length of an arc, $x_i = x_i(\tau)$, $i = 1, \cdots, n$, $0 \leqq \tau \leqq 1$, is defined *relative to each fixed value of* t as the value of the integral,

$$\int_0^1 f[x(\tau), x'(\tau), t]d\tau,$$

which exists if the arc is sufficiently regular and is, moreover, independent of the parametrization on account of the assumed homogeneity of $f[x, \dot{x}, t]$. The *distance* between two points is the length of the shortest arc connecting the two points, and this also naturally will in general depend upon t even though the two points are fixed when t changes. The parametric equations of such a *geodesical* arc are well known to satisfy Euler's equations,

$$(2.1) \quad f_{x_i}[x(\tau), x'(\tau), t] - (d/d\tau)f_{\dot{x}_i}[x(\tau), x'(\tau), t] = 0 \quad i = 1, \cdots, n.$$

Concerning the system of differential equations,

$$(2.2) \qquad dx/dt = X^i(x_1, \cdots, x_n, t) = X^i[x, t], \qquad i = 1, \cdots, n,$$

we assume that the X are of class C'. Our two fundamental theorems may be stated as follows:

THEOREM 1. *Let us consider two integral arcs of the system* (2.2).

$C_0 : x_i = x_i(t, 0)$ and $C_1 : x_i = x_i(t, 1)$, $\quad i = 1, \cdots, n$, $t_0 \leqq t \leqq t + h$.

We let $D(t)$ *denote the distance relative to* t *between* $x(t,0)$ *and* $x(t,1)$ *measured along the geodesical arc* g_t. *We furthermore assume that all these* g_t *are included within an open region* R, *and that no complete geodesic obtained by extending a* g_t *contains entirely within* R *an arc whose end points are conjugate to each other* (in the sense of the calculus of variations).

If

$$(2.3)\quad \alpha(t) \leqq \{f_{x_i}[x,\lambda,t]X^i[x,t] + f_{\dot{x}_i}[x,\lambda,t]X^i{}_{x_j}[x,t]\lambda^j + f_t[x,\lambda,t]\} \leqq \beta(t)$$

for every set $\lambda^1,\cdots,\lambda^n$ $(=\lambda^1(t),\cdots,\lambda^n(t))$ *such that* $f[x,\lambda,t]=1$ *and at every point* $[x]$ *on* g_t, *then*

$$(2.4)\quad D(t_0)\exp\int_{t_0}^{t}\alpha(u)\,du \leqq D(t) \leqq D(t_0)\exp\int_{t_0}^{t}\beta(u)\,du, \qquad t_0 \leqq t \leqq +h.$$

THEOREM 2. *Let* C_0 *and* C_1 *be defined as in the preceding theorem. Let* Γ_{t_0} *be an arbitrary sufficiently regular curve, with parametric equations,* $x_i = \xi_i(\tau)$, $i = 1,\cdots,n$, $0 \leqq \tau \leqq 1$, *whose end points are at* $[x(t_0,0)]$ *and* $[x(t_0,1)]$ *respectively. Let* C_τ *be the curve whose parametric equations,* $x_i = x_i(t,\tau)$, *satisfy the system* (2.2) *for each fixed value of* τ *on the unit interval and take on the initial values* $x_i(t_0,\tau) = \xi_i(\tau)$. *Let* Γ_t *be the arc whose parametric equations, for each fixed* t, $(t_0 \leqq t \leqq t_0 + h)$, *are likewise* $x_i = x_i(t,\tau)$, $0 \leqq \tau \leqq 1$. *Let* $L(t)$ *denote the length of* Γ_t *relative to* t.

If

$$(2.3^*)\quad \alpha^*(t) \leqq \{f_{x_i}[x,\lambda,t]X^i[x,t] + f_{\dot{x}_i}[x,\lambda,t]X^i{}_{x_j}[x,t]\lambda^j + f_t[x,\lambda,t]\} \leqq \beta^*(t),$$

for every set of λ*'s such that* $f[x,\lambda,t]=1$ *and at every point* $[x]$ *of* Γ_t, *then*

$$(2.4^*)\quad L(t_0)\exp\int_{t_0}^{t}\alpha^*(u)\,du \leqq L(t) \leqq L(t_0)\exp\int_{t_0}^{t}\beta^*(w)\,du.$$

The proofs of these two theorems are obvious modifications of the proofs of the corresponding theorems in MP. Moreover Theorem 1 may be obtained as an easy consequence of Theorem 2 in the manner indicated on p. 299 of MP. For the sake of brevity we shall therefore prove here only Theorem 2. For this purpose, we observe that, by our definition of length relative to t,

$$(2.5)\qquad L(t) = \int_0^1 f[x(t,\tau), x_\tau(t,\tau), t]\,d\tau.$$

Differentiating under the integral sign, we have (after making use of (2. 2))

$$(2.\,6)\quad L'(t) = \int_0^1 \{f_t[x(t,\tau), x_\tau(t,\tau), t] + f_{x_i}[x(t,\tau), x_\tau(t,\tau), t]X^i[x(t,\tau), t] \\ + f_{\dot{x}_i}[x(t,\tau), x_\tau(t,\tau), t]X^i{}_{x_j}[x(t,\tau), t]\partial x_j/\partial\tau\}d\tau.$$

We introduce the arc length s as a new parameter defined by

$$(2.\,7)\qquad s = s(t,\tau) = \int_0^\tau f[x(t,\tau'), x_\tau(t,\tau'), t]d\tau',$$

which has the obvious properties

$$(2.\,8)\qquad \partial s/\partial\tau = f[x(t,\tau), x_\tau(t,\tau), t] > 0, \qquad s(t,1) = L(t).$$

Moreover, on account of the assumed homogeneity of $f[x, \dot{x}, t]$, the parameter in (2. 7) can be changed from τ' to s', where $s' = s(t,\tau')$, with the result, deduced from (2. 7), that

$$s = \int_0^s f[x\{t, s'\}, x_s\{t, s'\}, t]ds',$$

where $x\{t, s(t,\tau)\} = x(t,\tau)$, so that

$$(2.\,9)\qquad f[x\{t,s\}, x_s\{t,s\}, t] = ds/ds = 1.$$

The integrand in (2. 6) is positively homogeneous of degree $+1$ in $\dot{x}_1, \cdots, \dot{x}_n$, because $f[x, \dot{x}, t]$ is. Hence (2. 6) may be written in the form

$$L'(t) = \int_0^{L(t)} \{f_t[x, x_s, t] + f_{x_i}[x, x_s, t]X^i[x, t] \\ + f_{\dot{x}_i}[x, x_s, t]X^i{}_{x_j}[x, t](\partial x_j/\partial s)\}ds.$$

It now follows at once from (2. 3*), by taking $\lambda^i = x^i{}_s$ in virtue of (2. 9), that

$$\alpha^*(t)L(t) \leqq L'(t) \leqq \beta^*(t)L(t),$$

from which the proof of Theorem 2 is completed by a simple integration.

3. Special case of a Riemann metric. Second order linear equations. Of particular importance is the case when

$$(3.\,1)\qquad f[x, \dot{x}, t] = (g_{ij}[x,t]\dot{x}_i\dot{x}_j)^{\frac{1}{2}}.$$

The middle member of (2. 3), or (2. 3*), we now verify can be written as follows:

$$(3.\,2)\qquad Q^*[\lambda] = (\tfrac{1}{2}X^k\partial g_{ij}/\partial x_k + \tfrac{1}{2}\partial g_{ij}/\partial t + g_{ik}X^k{}_{x_j})\lambda^i\lambda^j.$$

In obtaining this result, we of course make use of the fact that $\lambda^1, \cdots, \lambda^n$ is a unit vector, i. e. $g_{ij}\lambda^i\lambda^j = 1$. The $\alpha(t)$ and $\beta(t)$ of the previous section are therefore lower and opper bounds respectively of the quadratic form $Q^*[\lambda]$ for values of the λ's which are components of unit vectors.

The power of our fundamental theorems of the previous section, when a Riemann metric is employed, can be illustrated in the case of second order linear systems as follows:

Consider the second order linear differential equation

$$(px')' + qpx' + rx = 0 \tag{3. 3}$$

in which the primes denote differentiation with respect to the independent variable t, and in which $p, q, r, (pr)', (pq)'$ are continuous functions of t for $t_0 \leqq t < T$. By setting $x_1 = x$ and $x_2 = px'$, we may write the equivalent system in the form,

$$dx_1/dt = p^{-1}x_2 = X^1, \qquad dx_2/dt = -rx_1 - qx_2 = X^2, \tag{3. 4}$$

which can evidently be considered as a special case of (2. 2) with $n = 2$. With this system, assuming $pr - (1/4)(pq)^2 > 0$, we consider the Riemann metric

$$g_{11} = pr, \qquad g_{12} = g_{21} = \tfrac{1}{2}pq, \qquad g_{22} = 1, \tag{3. 5}$$

which, after a simple calculation and upon setting $\lambda^1 = \lambda$, $\lambda^2 = \mu$, allows us to specialize (3. 2) in the form,

$$Q^*[\lambda] = \tfrac{1}{2}q\{pr\lambda^2 + pq\lambda\mu + \mu^2\} + \tfrac{1}{2}\{(pr'\lambda^2 + (pq)'\lambda\mu\}$$

Since (λ, μ), as a unit vector, must satisfy the relation $pr\lambda^2 + pq\lambda\mu + \mu^2 = 1$, it is clear that

$$Q^*[\lambda] = -\tfrac{1}{2}q + \tfrac{1}{2}(pr)'\lambda^2 + \tfrac{1}{2}(pq)'\lambda\mu. \tag{3. 6}$$

Moreover, from the identities,

$$\begin{aligned} pr\lambda^2 + pq\lambda\mu + \mu^2 &\equiv (pr - \tfrac{1}{4}p^2q^2)\lambda^2 + (\tfrac{1}{2}pq\lambda + \mu)^2 \\ &\equiv ([pr]^{\frac{1}{2}}\lambda \pm \mu)^2 + (pq \mp 2[pr]^{\frac{1}{2}})\lambda\mu, \end{aligned}$$

it is readily deduced that our unit vector components must satisfy the inequalities,

$$\lambda^2 \leqq (pr - \tfrac{1}{4}p^2q^2)^{-1}, \tag{3. 7}$$

$$-(2[pr]^{\frac{1}{2}} - pq)^{-1} \leqq \lambda\mu \leqq (2[pr]^{\frac{1}{2}} + pq)^{-1}. \tag{3. 8}$$

Thus from (3. 6), (3. 7), and (3. 8), it is very easy to obtain explicit lower and upper bounds for $Q[\lambda]$. Let us consider somewhat further two special cases:

Case 1. $q \equiv 0$, $p > 0$, $r > 0$, and $(pr)' \geqq 0$. This is the self-adjoint case considered for instance by Wintner[2] and Leighton.[3]

Here we clearly have $\alpha(t) \equiv 0 \leqq Q^*[\lambda] \leqq \frac{1}{2}(pr)'/(pr) = \beta(t)$. Hence from Theorem 1, we find that $D(t_0) \leqq D(t) \leqq D(t_0)(pr/p_0r_0)^{\frac{1}{2}}$, where p_0 and q_0 are the values of p and r respectively when $t = t_0$. Here, of course, $D(t)$ represents the distance between two arbitrary solutions with respect to the metric appropriate to the value t of the independent variable. The geodesics are straight lines, since the Riemann metric, though dependent upon t, is independent of x_1 and x_2. Thus, if we compare the distance between an arbitrary solution $(x_1(t), x_2(t))$ of (3. 4) and the trivial solution $(0, 0)$, we see that (3. 7) yields upper and lower bounds for the quantity $D(t) = (pr(x_1(t))^2 + (x_2(t))^2)^{\frac{1}{2}}$. This shows a fortiori that

$$(pr)^{\frac{1}{2}} \,|\, x_1(t) \,|\, \leqq D(t_0)(pr/p_0r_0)^{\frac{1}{2}}.$$

Hence, dividing by $(pr)^{\frac{1}{2}}$, we come out with the result of Wintner and Leighton, previously referred to, which states that every solution of (3. 3) is bounded, if $q \equiv 0$, $p > 0$, $r > 0$, and $(pr)' \geqq 0$.

Case 2. $p \equiv 1$, $\lim\limits_{t \to \infty} q(t) = 0$, $\lim\limits_{t \to \infty} r(t) = R$, where R is a positive constant. We also assume that

$$\int_{t_0}^{+\infty} | q'(t) | \, dt \text{ and } \int_{t_0}^{\infty} | r'(t) | \, dt$$

both converge for a suitably chosen value of t_0. From these special conditions, we easily see that the general condition, $pr - (1/4)(pq)^2 > 0$, is satisfied for $t \geqq t_0$, if t_0 is large enough; and, from (3. 7) and (3. 8), we see that both λ^2 and $| \lambda\mu |$ are bounded, say, $\lambda^2 \leqq A$, $| \lambda\mu | \leqq A$. Hence, from (3. 6), we are justified in writing

$$\begin{aligned} \alpha(t) &= -\tfrac{1}{2}q(t) - \tfrac{1}{2}A(| q'(t) | + | r'(t) |) \\ &\qquad \leqq Q^*[\lambda] \leqq -\tfrac{1}{2}q(t) + \tfrac{1}{2}A(| q'(t) | + | r'(t) |) = \beta(t). \end{aligned}$$

Thus, from Theorem 1, we find that

$$D(t_0)K^{-1} \exp(-\tfrac{1}{2}\int_{t_0}^{t} q(u)\,du) \leqq D(t) \leqq D(t_0)K \exp(-\tfrac{1}{2}\int_{t_0}^{t} q(u)\,dw),$$

where $K = \exp(+\frac{1}{2}A \int_{t_0}^{\infty} (|r'(t)| + | q'(t) |)\,dt)$.

Here $D(t)$, as in Case 1, may be taken to represent the distance between an arbitrary solution $(x_1(t), x_2(t))$ of (3.4) and the trivial solution $(0, 0)$, so that

$$D(t) = (rx_1^2 + qx_1x_2 + x_2^2)^{\frac{1}{2}} = [(r - \tfrac{1}{4}q^2)x_1^2 + (\tfrac{1}{2}qx_1 + x_2)^2]^{\frac{1}{2}}$$

It thus follows, for example, that

$$|x_1(t)| \leqq (r - \tfrac{1}{4}q^2)^{-\frac{1}{2}}KD(t_0)\exp(-\tfrac{1}{2}\int_{t_0}^{t} q(u)\,du)$$
$$\leqq K^*D(t_0)\exp(-\tfrac{1}{2}\int_{t_0}^{t} q(u)\,du),$$

where K^* is a new constant determined in an obvious enough way.

Notice, for example, that the Bessel differential equation can be considered under both Case 1 and Case 2. Under Case 1, we write it in the form $(tx')' + (t - n^2t^{-1})x = 0$, taking $p = t$, $q = 0$, $r = t - n^2/t$, and we conclude that every solution is bounded for $t > t_0 > 0$. But, if we consider it under Case 2, by writing it in the form $x'' + t^{-1}x' + (1 - n^2t^{-2})x = 0$, taking $p = 1$, $q = t^{-1}$, and $r = (1 - n^2t^{-2})$, we obtain the familiar sharper result that $x(t) = O(t^{-\frac{1}{2}})$.

4. Approximation and linearization in connection with the present metric theory. The effectiveness of any application of our theory depends upon the choice of the metric, that is of the functions $g_{ij}[x, t]$. Very little can be said in a completely general way as to the most efficient way of choosing this metric. In some important cases, however, the key to the situation can be found by a study of the linear case, as we explain more fully later on.

Suppose now that upon making certain approximations a system (2.2) reduces to a simpler system,

$$dx_i/dt = L^i[x, t], \qquad i = 1, \cdots, n. \tag{4.1}$$

Then, if the approximations are justified, it is clear that a metric suited for the study of (4.1) will also be suited for the study of (2.2). Moreover the validity of our approximations can be appraised with the help of this same metric, $g_{ij}dx_idx_j$. Namely we can introduce an extra variable x_{n+1} and consider the $(n+1)$-*th* order system

$$\begin{aligned} dx_i/dt &= x_{n+1}L^i[x, t] + (1 - x_{n+1})X^i[x, t], \qquad i = 1, \cdots, n; \\ dx_{n+1}/dt &= 0 \end{aligned} \tag{4.2}$$

in connection with the $(n+1)$ dimensional metric, $\sum_{i,j=1}^{n} g_{ij}dx_idx_j + dx^2_{n+1}$.

Any solution of (4.2) for which $x_{n+1} = 1$ yields a solution of (4.1) and any solution of (4.2) for which $x_{n+1} = 0$ yields a solution of (2.2). Hence our Theorem 1 may be applied to compare those solutions of (2.2) and (4.1) which have the same initial values. These may be thought of as two solutions of the $(n+1)$-th order system (4.2) which initially are at a unit's distance apart in the $(n+1)$ dimensional space. One solution lies on the integral manifold $x_{n+1} = 0$, the other on $x_{n+1} = 1$. The length of any arc joining a point on one of these manifolds to a point on the other manifold must be $\geqq 1$ and can be equal to 1 only if the $x_1, \cdots, x_n$ coordinates of both points are the same. This follows from the fact that $\sum_{i,j=1}^{n} g_{ij}x_ix_j$ is positive definite. Hence a rigorous measure of an upper bound for the deviation of the solution of the "approximate" equations (4.1) from the solution of the "exact" equations (2.2) could be taken to be $D(t) - 1$ or $(D(t)^2 - 1)^{\frac{1}{2}}$, where $D(t) \leqq \exp \int_{t_0}^{t} \beta(w)\,du$. Here, of course, $D(t)$ and β are all formed with reference to the $(n+1)$ dimensional problem, and $\beta(t)$ is an upper bound of a certain quadratic form in the components of an $(n+1)$ dimensional unit vector (cf. (3.2)).

Let us now consider equations (2.2) together with approximations leading to equations (4.1) which are supposed to split naturally into three groups. First, we have a set of μ equations involving only the unknowns, $x_1, \cdots, x_\mu$:

$$(4.3) \qquad dx_i/dt = L^i(x_1, \cdots, x_\mu, t)\,; \qquad i = 1, 2, \cdots, \mu.$$

Secondly we have a set of ν equations which are linear in the ν unknowns $x_{\mu+1}, \cdots, x_{\mu+\nu}$, with coefficients which are functions of $x_1, \cdots, x_\mu$, and t:

$$(4.4) \qquad dx_i/dt = L^i(x_1, \cdots, x_{\mu+\nu}, t)$$
$$= \sum_{j=\mu+1}^{\mu+\nu} a^i{}_j(x_1, \cdots, x_\mu, t)x_j + b^i(x_1, \cdots, x_\mu, t)$$
$$i = \mu + 1, \cdots, \mu + \nu.$$

Thirdly we have the remaining $\rho = n - (\mu + \nu)$ equations of completely general type:

$$(4.5) \qquad dx_i/dt = L^i[x, t], \qquad i = \mu + \nu + 1, \cdots, \mu + \nu + \rho = n.$$

Here, we asume that $\rho \geqq 0$, and we need not exclude the vacuous case $\rho = 0$.

Such a situation arises in exterior ballistics, when for example we make a linearizing approximation by assuming known functions of time for the axial velocity, axial spin, and direction parameters for the shell axis. Denoting these quantities by x_1, x_2, x_3, x_4, we can formulate these assumptions in the form of $\mu = 4$ differential equations of the form (4. 3) in which the right hand members depend simply on t. In such cases, of course, the integration of (4. 3) is immediately effected by quadratures. Upon substituting the resulting values of $x_1, \cdots, x_\mu$ in (4. 4), we obtain a linear system for $x_{\mu+1}, \cdots, x_{\mu+\nu}$ and it is here that the most interesting problems of stability arise in connection with the approximate ballistic equations, in which $\nu = 4$ and in which x_5, x_6, x_7, x_8 denote the cross components of velocity and spin. If first (4. 3) and then (4. 4) can be solved, the approximate solution is completed by solving (4. 5), an independent system of order ρ. In the ballistic case this can be done by quadratures. On account of the central importance of the linear system (4. 4) we shall therefore devote the following section to a discussion of a favorable metric for its study.

The metric so obtained will depend upon the coefficients of the linear system which in turn depend upon the particular solution of (4. 2), which has been inserted into these coefficients. In other words the metric of the linear system will be of the form,

$$\sum_{i,j=\mu+1}^{\mu+\nu} g_{ij}(x_1, \cdots, x_\mu, t)\, dx_i dx_j,$$

and then it is natural to set up the metric for the entire system (4. 1), and thus also for (2. 2) in the form,

$$\sum_{k=1}^{\mu} A_k dx_k^2 + \sum_{i,j=\mu+1}^{\mu+\nu} g_{ij}(x_1, \cdots, x_\mu, t)\, dx_i dx_j + \sum_{h=\mu+\nu+1}^{n} B_h dx_h^2,$$

where the A's and B's are merely positive constants chosen for convenience in weighting the importance of the different dependent variables. We cannot discuss here the choice of these constants.

5. The metric for general linear systems. In discussing the metric for the linear system (4. 4) we change our notation writing it in the form $dx_i/dt = \sum_{i,j=1}^{n} a_{ij}(t)x_j + b_i(t)$. In applications, of course, all derivatives of the a's with respect to t must be taken in the sense $\sum_{k=1}^{\mu} (\partial a/\partial x_k)L^k[x, t] + \partial a/\partial t$, if, as in (4. 4), the a's depend on t not only explicitly but through the mediation of $x_1, \cdots, x_\mu$.

THEOREM 3. *Given any non-singular real matrix A with simple elemen-*

tary divisors, it is always possible to find a positive definite symmetric matrix G *(of the same order as* A*) such that the roots* $\lambda_1, \cdots, \lambda_n$ *of the equation,*

$$\det\left((B - 2\lambda G\right) = 0, \quad \textit{where } B = GA + A'G,$$

are precisely the real parts of the characteristic roots of A. If the elements of A are functions of a variable t, it is of course necessary to assume that A has simple elementary divisors for all values of t for which $G = G(t)$ is required.

The following proof gives a construction of G:

Proof. If the complex characteristic roots of A be denoted by $\alpha_k \pm i\beta_k$, $k = 1, 2, \cdots, s$, the transformation $y = Ax$ (where x and y represent vectors $x_1, \cdots, x_n$, and $y_1, \cdots, y_n$, respectively) can be written in the canonical form

$$\eta_{2k-1} = \alpha_k\xi_{2k-1} - \beta_k\xi_{2k}, \quad \eta_{2k} = \beta_k\xi_{2k-1} + \alpha_k\xi_{2k}, \qquad k = 1, \cdots, s,$$
$$\eta_h = \gamma_h\xi_h, \qquad h = 2s + 1, \cdots, n,$$

when ξ and η are real vectors related to x and y by means of equations of the form, $\xi = Cx$, $\eta = Cy$, $\det C \neq 0$.
Let

$$Q = \sum_{j=1}^{n} \xi_j^2 = \sum_{j=1}^{n} \Big(\sum_{i=1}^{n} c_{ji}x_i \Big)^2 = \sum_{i,j} g_{ij}x_ix_j, \quad \text{where } g_{ij} = \sum_{h=1}^{n} c_{hi}c_{hj} = g_{ji}.$$

We regard Q as an invariant with respect to transformations of coordinates. Hence the vector $\partial Q/\partial x_1, \cdots, \partial Q/\partial x_n$ is covariant, while the vector $y_1, \cdots, y_n$ can be regarded as contravariant (in spite of the employment of subscripts instead of the more usual superscripts for contravariant vectors). It follows from tensor analysis that $\bar{Q} = \sum_{i=1}^{n} y_i \partial Q/\partial x_i$ is also an invariant and hence that

$$\bar{Q} = \sum_{h=1}^{n} \eta_h \cdot \partial Q/\partial\xi_h = \sum_{k=1}^{s} 2\alpha_k(\xi^2_{2k-1} + \xi^2_{2k}) + \sum_{h=2s+1}^{n} 2\gamma\xi^2_h.$$

We see immediately, from this representation of the quadratic form $\bar{Q}$, that the stationary values of $\bar{Q}$ under the condition $Q = 1$ are precisely the real numbers $2\alpha_k$, $2\gamma_h$. But these stationary values are also known to be the roots $\sigma_1, \cdots, \sigma_n$ of the equation $\det(B - \sigma G) = 0$ where $G = (g_{ij})$. The theorem follows at once when we replace σ by 2λ.

The significance of Theorem 3 becomes obvious when we consider a system of linear differential equations of the form, $dx_i/dt = \sum_j a_{ij}(t)x_j$. If $Q = \sum g_{ij}x_ix_j$ and $\dot{Q} = \sum \dot{g}_{ij}x_ix_j$ $(\dot{g}_{ij} = dg_{ij}/dt)$, we estimate the rate of growth of Q as follows:

$$dQ/dt = 2 \sum g_{ij} x_i dx_j/dt + \sum \dot{g}_{ij} x_i x_j = 2 \sum_{i,j,k} g_{ij} a_{jk} x_i x_k + \sum_{i,j} \dot{g}_{ij} x_i x_j = \bar{Q} + \dot{Q},$$

where $\bar{Q}$, as above, is the quadratic form whose coefficients form the matrix $GA + A'G$. Hence

$$(5.1) \qquad (dQ/dt)/Q = (\bar{Q} + \dot{Q})/Q$$

is less than or equal to the maximum value of $\bar{Q} + \dot{Q}$ under the condition $Q = 1$. This follows from the fact that the right member of (5.1) is homogeneous of degree 0 in $x_1, \cdots, x_n$. This is true no matter how G is chosen, so long as it is positive definite. If, however, it is chosen in accordance with Theorem 3 and if the derivatives $\dot{a}_{ij}$ and hence $\dot{g}_{ij}$ are small (as in the ballistic case), we can say that the *maximum* value of $d \log Q/dt$ is not greater than twice the algebraically greatest real part of the characteristic roots of A increased by a small quantity, namely the maximum of the form $\dot{Q}$ under the condition $Q = 1$. A similar statement may be made regarding the minimum value of $d \log Q/dt$.

It is easily verified, as one would expect, that, with our present $g_{ij}(t)$ taken as the coefficients of a Riemann metric depending on t but independent of $x_1, \cdots, x_n$, and with $X^i[x, t] = \sum_j a_{ij}(t) x_j + b_i(t)$, the $Q^*[\lambda]$ of Section **3** is here the same as one half the quadratic form $\bar{Q} + \dot{Q}$ with the x's replaced by λ's.

In the examples of Section **3**, the matrix G, that is, the metric $g_{ij} dx_i dx_j$, was indeed chosen in accordance with Theorem 3. The two characteristic roots in each example were complex with real parts 0, in the first example (Case 1) and $\frac{1}{2}q$ in the second example (Case 2). Thus, it is seen in more detail that our fundamental Theorem 1 is capable of generalizing the rather refined results regarding these two examples to general linear systems in which the coefficient matrix has simple elementary divisors.

THE JOHNS HOPKINS UNIVERSITY,
NAVAL ORDNANCE LABORATORY, WHITE OAK, MARYLAND.

THE m-th POWER RESIDUE SYMBOL.*

By W. H. Mills.

---

1. Introduction. Let k be an algebraic number field containing the primitive m-th roots of unity. The m-th power residue symbol has been widely studied and many of its properties are known. The quotient $\left(\frac{\alpha}{\beta}\right)\left(\frac{\beta}{\alpha}\right)^{-1}$ has received much attention and has been evaluated in a number of special cases. In all these investigations the field k has been held fixed. The object of this paper is to investigate the behavior of the power residue symbols when the field k and the exponent m are varied.

2. Definitions. Let m be a positive integer and k an algebraic number field containing a primitive m-th root of unity ζ. Let $\mathfrak{p}$ be a prime ideal of k not dividing m.

Lemma 1. *$\zeta^a \equiv \zeta^b \pmod{\mathfrak{p}}$ if and only if $\zeta^a = \zeta^b$.*

Proof.

$$\prod_{\nu=1}^{m-1}(x-\zeta^\nu) = x^{m-1}+x^{m-2}+\cdots+1. \quad \text{If } x=1 \text{ then } \prod_{\nu=1}^{m-1}(1-\zeta^\nu) = m.$$

Suppose $\zeta^a \equiv \zeta^b \pmod{\mathfrak{p}}$. Then $\mathfrak{p} \mid 1-\zeta^{a-b}$. It follows that $\zeta^{a-b}=1$, and therefore $\zeta^a = \zeta^b$. The converse is immediate.

Let α be an element of k relatively prime to $\mathfrak{p}$. Then $\alpha^{N\mathfrak{p}-1} \equiv 1 \pmod{\mathfrak{p}}$. In particular $\zeta^{N\mathfrak{p}-1} \equiv 1 \pmod{\mathfrak{p}}$, and so by Lemma 1, $\zeta^{N\mathfrak{p}-1} = 1$. Therefore $m \mid N\mathfrak{p}-1$. Now $(\alpha^{(N\mathfrak{p}-1)/m})^m \equiv 1 \pmod{\mathfrak{p}}$. The equation $x^m \equiv 1 \pmod{\mathfrak{p}}$ has exactly m solutions, namely $1, \zeta, \zeta^2, \cdots, \zeta^{m-1}$. Therefore $\alpha^{(N\mathfrak{p}-1)/m} \equiv \zeta^r \pmod{\mathfrak{p}}$ for some r.

Definition 1. $\left(\frac{\alpha}{\mathfrak{p}}\right)_{km}$ is the unique m-th root of unity satisfying

$$\left(\frac{\alpha}{\mathfrak{p}}\right)_{km} \equiv \alpha^{(N\mathfrak{p}-1)/m} \pmod{\mathfrak{p}}.$$

Definition 2. Let A be a non-zero field element or ideal of k, relatively prime to α and m. Then

$$\left(\frac{\alpha}{A}\right)_{km} = \prod_{\mathfrak{p}}\left(\frac{\alpha}{\mathfrak{p}}\right)_{km}^{\operatorname{ord}_{\mathfrak{p}} A},$$

---

* Received January 25, 1950.

where the product is taken over all prime ideals of k that are relatively prime to α and m.

The subscripts k and m will be dropped when it can be done without danger of ambiguity.

It is easily shown that $\left(\frac{\alpha}{\mathfrak{p}}\right) = 1$ if and only if α is congruent to an m-th power modulo $\mathfrak{p}$. For this reason the symbol $\left(\frac{\alpha}{\mathrm{A}}\right)$ is called an m-th power residue symbol. It is to be noted that it is defined if and only if k is an algebraic number field containing the primitive m-th roots of unity, and A is a non-zero element or ideal of k relatively prime to m and α. It will be assumed in the following discussion that all residue symbols occurring in the statements of lemmas and theorems are defined.

3. Properties of the power residue symbol. From Definitions 1 and 2 it follows immediately that $\left(\frac{\alpha}{\mathrm{A}}\right)$ is multiplicative in both α and A. Thus we have:

LEMMA 2. *The m-th power residue symbol is doubly multiplicative, i. e.*

$$\left(\frac{\alpha}{\mathrm{A}}\right)\cdot\left(\frac{\beta}{\mathrm{A}}\right)=\left(\frac{\alpha\beta}{\mathrm{A}}\right) \quad \textit{and} \quad \left(\frac{\alpha}{\mathrm{A}}\right)\cdot\left(\frac{\alpha}{\mathrm{B}}\right)=\left(\frac{\alpha}{\mathrm{AB}}\right).$$

We will now investigate how the value of the power residue symbol varies when the field k and the power m are varied with α and A held fixed. Because of the multiplicative properties of the residue symbol it will be sufficient to prove the following lemmas only for the case $\mathrm{A} = \mathfrak{p}$, a prime ideal of k. It follows from Lemma 1 that to prove two residue symbols equal, say an m-th power and an n-th power residue symbol, it is sufficient to prove them congruent modulo a prime ideal not dividing mn.

LEMMA 3. *Let K be a finite extension of k with degree n. Then*

$$\left(\frac{\alpha}{\mathrm{A}}\right)_{Km}=\left(\frac{\alpha}{\mathrm{A}}\right)_{km}^{n}.$$

Proof. Without loss of generality let $\mathrm{A} = \mathfrak{p}$, a prime ideal of k. Then $m \mid N\mathfrak{p} - 1$, so we may put $N\mathfrak{p} = 1 + bm$, b a rational integer. Let $\mathfrak{P}$ be a prime ideal of K such that $\mathfrak{P} \mid \mathfrak{p}$ with residue class field degree f.

$$(N\mathfrak{P} - 1)/m = (N\mathfrak{p}^{f} - 1)/m = ((1 + bm)^{f} - 1)/m \equiv bf = f(N\mathfrak{p} - 1)/m \pmod{bm}.$$

Since $\alpha^{bm} = \alpha^{N\mathfrak{p}-1} \equiv 1 \pmod{\mathfrak{p}}$, it follows that

$$\left(\frac{\alpha}{\mathfrak{P}}\right)_{Km} \equiv \alpha^{(N\mathfrak{P}-1)/m} \equiv \alpha^{f(N\mathfrak{p}-1)/m} \equiv \left(\frac{\alpha}{\mathfrak{p}}\right)_{km}^{f} \pmod{\mathfrak{P}}.$$

Now $\mathfrak{P} \nmid m$. Therefore $\left(\frac{\alpha}{\mathfrak{P}}\right)_{Km} = \left(\frac{\alpha}{\mathfrak{p}}\right)_{km}^{f}$. Let $\mathfrak{p} = \prod \mathfrak{P}_\nu^{e_\nu}$ be the factorization of $\mathfrak{p}$ in K. Then

$$\left(\frac{\alpha}{\mathfrak{p}}\right)_{Km} = \prod \left(\frac{\alpha}{\mathfrak{P}_\nu}\right)_{Km}^{e_\nu} = \prod \left(\frac{\alpha}{\mathfrak{p}}\right)_{km}^{e_\nu f_\nu} = \left(\frac{\alpha}{\mathfrak{p}}\right)_{km}^{\Sigma e_\nu f_\nu} = \left(\frac{\alpha}{\mathfrak{p}}\right)_{km}^{n}.$$

LEMMA 4. *Let σ be an automorphism of k. Then* $\sigma\left(\frac{\alpha}{\mathrm{A}}\right) = \left(\frac{\sigma\alpha}{\sigma\mathrm{A}}\right)$.

Proof. Without loss of generality let $\mathrm{A} = \mathfrak{p}$. By definition

$$\alpha^{(N\mathfrak{p}-1)/m} \equiv \left(\frac{\alpha}{\mathfrak{p}}\right) \quad (\mathrm{mod}\ \mathfrak{p}), \text{ hence } (\sigma\alpha)^{(N\mathfrak{p}-1)/m} \equiv \sigma\left(\frac{\alpha}{\mathfrak{p}}\right) \quad (\mathrm{mod}\ \sigma\mathfrak{p}),$$

and since $N(\sigma\mathfrak{p}) = N\mathfrak{p}$ we have

$$\left(\frac{\sigma\alpha}{\sigma\mathfrak{p}}\right) \equiv (\sigma\alpha)^{(N\sigma\mathfrak{p}-1)/m} \equiv \sigma\left(\frac{\alpha}{\mathfrak{p}}\right) \quad (\mathrm{mod}\ \sigma\mathfrak{p}).$$

Lemma 4 follows immediately.

LEMMA 5. *Let $\alpha \varepsilon k$, A be a non-zero element or ideal of k, and let K be a finite extension of k containing primitive n-th roots of unity. Then* $\left(\frac{\alpha}{\mathrm{A}}\right)_{Kn} \varepsilon\, k$.

Proof. Put $L = k(\epsilon)$, where ϵ is a primitive n-th root of unity. Then $K \supseteq L$, and L is a normal extension of k. Let σ be any automorphism of L/k. Using Lemma 4 we obtain

$$\sigma\left(\frac{\alpha}{\mathrm{A}}\right)_{Ln} = \left(\frac{\sigma\alpha}{\sigma\mathrm{A}}\right)_{Ln} = \left(\frac{\alpha}{\mathrm{A}}\right)_{Ln}.$$

Therefore $\left(\frac{\alpha}{\mathrm{A}}\right)_{Ln} \varepsilon\, k$. By Lemma 3 we have $\left(\frac{\alpha}{\mathrm{A}}\right)_{Kn} = \left(\frac{\alpha}{\mathrm{A}}\right)_{Ln}^{(K:L)} \varepsilon\, k$.

LEMMA 6. $\left(\frac{\alpha}{\mathrm{A}}\right)_{k,mn}^{m} = \left(\frac{\alpha}{\mathrm{A}}\right)_{kn}$.

Proof. Without loss of generality let $\mathrm{A} = \mathfrak{p}$. Then

$$\left(\frac{\alpha}{\mathfrak{p}}\right)_{k,mn}^{m} \equiv \alpha^{(N\mathfrak{p}-1)/n} \equiv \left(\frac{\alpha}{\mathfrak{p}}\right)_{kn} \quad (\mathrm{mod}\ \mathfrak{p}).$$

Let $m = \prod l_\nu^{n_\nu}$ be the factorization of m into powers of distinct rational primes. Let a_ν be rational integers satisfying $\sum a_\nu m l_\nu^{-n_\nu} = 1$. Then by Lemma 6

$$\left(\frac{\alpha}{\mathrm{A}}\right)_{km} = \left(\frac{\alpha}{\mathrm{A}}\right)_{km}^{\Sigma a_\nu m l_\nu^{-n_\nu}} = \prod \left(\frac{\alpha}{\mathrm{A}}\right)_{k l_\nu^{n_\nu}}^{a_\nu}.$$

Thus the study of the m-th power residue symbol can be reduced to the study of prime power residue symbols. Furthermore if $r < n$ then by Lemma

6 we can express the l^r-th power residue symbol in terms of the l^n-th power symbol. Therefore it is sufficient to consider the l^n-th power residue symbol, where l is a prime and k does not contain primitive l^{n+1}-st roots of unity. When k contains the primitive l^n-th roots of unity but not the primitive l^{n+1}-st roots of unity we will say that k contains the l^n-th roots of unity exactly.

Let k contain the l^n-th roots of unity exactly, and let K be a finite extension of k containing ζ, a primitive l^{n+s} root of unity, $s \geqq 1$. We wish to express $\left(\frac{\alpha}{\mathrm{A}}\right)_{K\,l^{n+s}}$ in terms of $\left(\frac{\alpha}{\mathrm{A}}\right)_{k\,l^n}$. By Lemma 3 it is sufficient to consider the case $K = k(\zeta)$.

THEOREM 1. *Let l be a prime and n a non-negative integer. Let k contain the l^n-th roots of unity exactly, and put $K = k(\zeta)$, where ζ is a primitive l^{n+s} root of unity, $s \geqq 1$. Let R be the field of rational numbers. Then*

$$\left(\frac{\alpha}{\mathrm{A}}\right)_{K\,l^{n+s}} = \left(\frac{\alpha}{\mathrm{A}}\right)_{k\,l^n}^{E},$$

$$\textit{where } E = \begin{cases} 1 & \textit{if } l > 2, \\ 1 + 2^{n-1} & \textit{if } l = 2,\ n > 1, \\ 2 & \textit{if } l = 2,\ n = 1, \textit{ and } k \cap R(\zeta) \textit{ is a real field,} \\ 1 & \textit{if } l = 2,\ n = 1, \textit{ and } k \cap R(\zeta) \textit{ is a complex field.} \end{cases}$$

Proof. Case I. $(K:k) = l$. Let G be the Galois group of K/k. $N_{K/k}\zeta = \prod_{\sigma \varepsilon G} \sigma\zeta = N_{R(\zeta)/k \cap R(\zeta)}\zeta$ will be evaluated.

(1) $l^n > 2$. Here $s = 1$ and

$$N_{K/k}\zeta = \prod_{\sigma \varepsilon G} \sigma\zeta = \prod_{\nu=0}^{l-1} \zeta^{1+\nu l^n} = \zeta^{l + l^{n+1}(l-1)/2} = \zeta^{l^s E}.$$

(2) $l^n = 1$. $N_{K/k}\zeta$ is an l^s root of unity contained in k, but k does not contain primitive l-th roots of unity. Therefore $N_{K/k}\zeta = 1 = \zeta^{l^s E}$. We note that the case $l = 2$, $n = 0$ is excluded by hypothesis.

(3) $l^n = 2$, $k \cap R(\zeta)$ is real. $N_{K/k}\zeta$ is a positive real number and a root of unity. Therefore $N_{K/k}\zeta = 1 = \zeta^{2 l^s} = \zeta^{l^s E}$.

(4) $l^n = 2$, $k \cap R(\zeta)$ is complex. The field $R(\zeta)$ has three subfields of degree 2, the field of the 2^{n+s-1}-st roots of unity which is fixed under the automorphism $\zeta \to -\zeta$, a real field fixed under $\zeta \to 1/\zeta$, and a complex field not containing i, namely $k \cap R(\zeta)$, which is fixed under the automorphism $\zeta \to -1/\zeta$. Thus $N_{K/k}\zeta = -1 = \zeta^{l^s E}$. Therefore $N_{K/k}\zeta = \zeta^{l^s E}$ in all cases. It follows immediately that if ϵ is any l^{n+s} root of unity then $N_{K/k}\epsilon = \epsilon^{l^s E}$. Without loss of generality let $\mathrm{A} = \mathfrak{p}$, a prime ideal of k. Let $\mathfrak{P}$ be a prime ideal of K dividing $\mathfrak{p}$ with residue class field degree f. Then $f = 1$ or $f = l$.

(A) Suppose $f=1$. Then $\mathfrak{p}=\prod_{\sigma\in G}\sigma\mathfrak{P}$, and

$$\left(\frac{\alpha}{\mathfrak{p}}\right)_{Kl^{n+s}}=\prod_{\sigma\in G}\left(\frac{\alpha}{\sigma\mathfrak{P}}\right)_{Kl^{n+s}}=\prod_{\sigma\in G}\sigma\left(\frac{\alpha}{\mathfrak{P}}\right)_{Kl^{n+s}}=\left(\frac{\alpha}{\mathfrak{P}}\right)^{l^sE}_{Kl^{n+s}}$$

$$\equiv\alpha^{(N\mathfrak{P}-1)\,l^{-n}E}=\alpha^{(N\mathfrak{p}-1)\,l^{-n}E}\equiv\left(\frac{\alpha}{\mathfrak{p}}\right)^{E}_{kl^n}\pmod{\mathfrak{P}}.$$

(B) Suppose $f=l$. Then

$$(N\mathfrak{P}-1)/(N\mathfrak{p}-1)=(N\mathfrak{p}^l-1)/(N\mathfrak{p}-1)=1+N\mathfrak{p}+\cdots+N\mathfrak{p}^{l-1}.$$

Therefore by the theory of finite fields

$$\zeta^{(N\mathfrak{P}-1)/(N\mathfrak{p}-1)}\equiv\prod_{\sigma\in G}\sigma\zeta=\zeta^{l^sE}\pmod{\mathfrak{p}}.$$

It follows that $(N\mathfrak{P}-1)/(N\mathfrak{p}-1)\equiv l^sE\pmod{l^{n+s}}$ or

$$(N\mathfrak{P}-1)l^{-s}/(N\mathfrak{p}-1)\equiv E\pmod{l^n}.$$

Therefore

$$\left(\frac{\alpha}{\mathfrak{p}}\right)_{Kl^{n+s}}=\left(\frac{\alpha}{\mathfrak{P}}\right)_{Kl^{n+s}}\equiv\alpha^{(N\mathfrak{P}-1)\,l^{-n-s}}$$

$$\equiv\left(\frac{\alpha}{\mathfrak{p}}\right)^{(N\mathfrak{P}-1)\,l^{-s}/(N\mathfrak{p}-1)}_{kl^n}=\left(\frac{\alpha}{\mathfrak{p}}\right)^{E}_{kl^n}\pmod{\mathfrak{P}}.$$

This completes the proof of Theorem 1 for the case $(K:k)=l$. Now $(K:k)$ is a positive power of l. There remains:

Case II. $(K:k)=l^r$, $r\geqq 2$. We suppose that the theorem holds for $(K:k)=l^{r-1}$ and for $(K:k)=l$. Now $(k(\zeta^l):k)=l^{r-1}$ and $k(\zeta^l)$ contains the l^{n+s-1}-st roots of unity exactly. Thus we have

$$\left(\frac{\alpha}{\mathrm{A}}\right)_{Kl^{n+s}}=\left(\frac{\alpha}{\mathrm{A}}\right)^{E'}_{k(\zeta^l),\,l^{n+s-1}}=\left(\frac{\alpha}{\mathrm{A}}\right)^{E'E}_{kl^n}$$

where $E'\equiv 1\pmod{l^{n+s-2}}$. Since $s\geqq r\geqq 2$ it follows that $E'E\equiv E\pmod{l^n}$. This completes the proof.

The result of Theorem 1 can be easily extended to arbitrary power residue symbols.

THEOREM 2. *Let K be a finite extension of the algebraic number field k. Let K contain the primitive r-th roots of unity. Let $r=\prod l_j^{s_j}$ be the factorization of r into powers of distinct rational primes, and let k contain the $l_j^{n_j}$ roots of unity exactly. Put $m=\prod l_j^{n_j}$. Let ζ_j be a primitive $l_j^{s_j}$-th root of unity. If $l_j^{n_j}>2$, or if $s_j\leqq n_j$, put $\Delta_j=0$; if $l_j=2$, $s_j>n_j=1$, and $\zeta_j\varepsilon k(i)$, put $\Delta_j=s_j-2$; and if $l_j=2$, $n_j=1$, and $\zeta_j\not\varepsilon k(i)$, let $k(i)$ contain the $2^{2+\Delta_j}$-th roots of unity exactly. Put $k_j=k(\zeta_j)$ and .*

$$E_j = \begin{cases} 1 & \textit{if } s_j \leqq n_j, \\ 1 & \textit{if } s_j > n_j,\ l_j > 2, \\ 1 + 2^{n-1} & \textit{if } l_j = 2,\ s_j > n_j > 1, \\ 2 & \textit{if } l_j = 2,\ s_j > n_j = 1, \textit{ and } k \cap R(\zeta_j) \textit{ is a real field}, \\ 1 & \textit{if } l_j = 2,\ s_j > n_j = 1, \textit{ and } k \cap R(\zeta_j) \textit{ is a complex field.} \end{cases}$$

Choose an integer Q satisfing

$$Qrl_j^{-s_j} \equiv E_j(K:k)ml_j^{\Delta_j - s_j} \qquad (\operatorname{mod} l_j^{n_j}) \tag{1}$$

for all j for which $n_j > 0$. Then

$$\left(\frac{\alpha}{\mathrm{A}}\right)_{Kr} = \left(\frac{\alpha}{\mathrm{A}}\right)_{km}^{Q} \tag{2}$$

for all α and A for which both sides are defined.

We note that the right side of the congruence (1) is integral if $n_j > 0$. Thus the integer Q is determined modulo m.

Proof. (I) Suppose $s_j > n_j > 0$. Then $(k_j : k) = l_j^{s_j - n_j - \Delta_j}$. By Lemmas 3 and 6 and Theorem 1 we have

$$\begin{aligned} \left(\frac{\alpha}{\mathrm{A}}\right)_{Kr}^{rl_j^{-s_j}} &= \left(\frac{\alpha}{\mathrm{A}}\right)_{Kl_j^{s_j}} = \left(\frac{\alpha}{\mathrm{A}}\right)_{k_j l_j^{s_j}}^{(K:k_j)} = \left(\frac{\alpha}{\mathrm{A}}\right)_{kl_j^{n_j}}^{E_j(K:k_j)} \\ &= \left(\frac{\alpha}{\mathrm{A}}\right)_{kl_j^{n_j}}^{E_j(K:k)\,l_j^{n_j+\Delta_j-s_j}} = \left(\frac{\alpha}{\mathrm{A}}\right)_{km}^{Qrl_j^{-s_j}}. \end{aligned} \tag{3}$$

(II) Suppose $s_j \leqq n_j$. Then $\Delta_j = 0$ and $E_j = 1$. By Lemmas 3 and 6

$$\left(\frac{\alpha}{\mathrm{A}}\right)_{Kr}^{rl_j^{-s_j}} = \left(\frac{\alpha}{\mathrm{A}}\right)_{kl_j^{s_j}}^{(K:k)} = \left(\frac{\alpha}{\mathrm{A}}\right)_{km}^{E_j(K:k)ml_j^{\Delta_j - s_j}} = \left(\frac{\alpha}{\mathrm{A}}\right)_{km}^{Qrl_j^{-s_j}} \tag{4}$$

(III) Suppose $n_j = 0$. Then $l_j \nmid m$ and k does not contain primitive l_j-th roots of unity. Whence

$$\left(\frac{\alpha}{\mathrm{A}}\right)_{Kr}^{mrl_j^{-s_j}} = \left(\frac{\alpha}{\mathrm{A}}\right)_{Kl_j^{s_j}}^{m} = 1 = \left(\frac{\alpha}{\mathrm{A}}\right)_{km}^{Qmrl_j^{-s_j}}. \tag{5}$$

The exponents appearing on the left side of (3), (4), and (5) are relatively prime. (2) follows easily.

COROLLARY. Under the hypothesis of Theorem 2, if $r \mid (K:k)m$ then

$$\left(\frac{\alpha}{\mathrm{A}}\right)_{Kr} = \left(\frac{\alpha}{\mathrm{A}}\right)_{km}^{(K:k)m/r} \text{ if } r \text{ is odd, and}$$

$$\left(\frac{\alpha}{\mathrm{A}}\right)_{Kr} = \left(\frac{\alpha}{\mathrm{A}}\right)_{km}^{(K:k)m(1+\frac{1}{2}m)/r} \text{ if } 4 \mid m.$$

YALE UNIVERSITY.

RECIPROCITY IN ALGEBRAIC NUMBER FIELDS.*

By W. H. MILLS.

1. Introduction. Let m be a positive rational integer and k be an algebraic number field containing primitive m-th roots of unity. Let α and β be elements of k that are relatively prime to each other and to m. The quotient of m-th power residue symbols, $\left(\frac{\alpha}{\beta}\right)\cdot\left(\frac{\beta}{\alpha}\right)^{-1}$, has been extensively studied[1] and many important properties of it are known. It is closely connected with the Hilbert norm residue symbol. Formulas for this quotient are known in many special cases, in particular for all cyclotomic fields and for certain extensions of cyclotomic fields.[2] The purpose of this paper is to obtain a general formula for this expression, valid for all algebraic number fields for which it is defined.[3]

2. The norm residue symbol. Let R be the field of rational numbers, and $\bar{R}$ be the completion of R at a prime p, not necessarily finite. Let $\bar{k}$ be a finite extension of $\bar{R}$ containing primitive m-th roots of unity. Let α and β be non-zero elements of $\bar{k}$, and $\mathfrak{p}$ be the prime of $\bar{k}$.

Definition. The Hilbert norm residue symbol $(\alpha, \beta, \bar{k}, m)$ is that m-th root of unity satisfying[4]

$$(\alpha, \beta, \bar{k}, m)\beta^{1/m} = (\alpha, \bar{k}(\beta^{1/m})/\bar{k} \mid \mathfrak{p})\beta^{1/m}.$$

The norm residue symbol will be written (α, β) when it can be done without danger of ambiguity. It has been frequently written $\left(\frac{\alpha, \beta}{\mathfrak{p}}\right)$.

The following properties of the norm residue symbol are well known.[5]

I. $(\alpha\beta, \gamma) = (\alpha, \gamma)(\beta, \gamma)$.

---

* Received January 25, 1950.

[1] See, for example, [4], part II.

[2] Rothgiesser [6] has obtained formulas that are applicable in any finite extension of the field of the m-th roots of unity that is unramified at all primes dividing m.

[3] In some recent papers, [7] and [8], Šafarevič has approached this problem using the p^n-th primary numbers introduced by Hasse [5]. His results are of a different nature from those of this paper.

[4] For details see [4], part II, p. 53.

[5] See [4], part II, pp. 54-59 and p. 75.

II. $(\alpha, \beta\gamma) = (\alpha, \beta)(\alpha, \gamma)$.

III. $(1 - \alpha, \alpha) = 1$.

IV. $(\beta, \alpha) = (\alpha, \beta)^{-1}$.

V. $(\alpha, \beta, \bar{k}, mn)^m = (\alpha, \beta, \bar{k}, n)$.

VI. Let α and β be non-zero elements of an algebraic number field k such that, if $\mathrm{ord}_{\mathfrak{q}}\, \alpha \neq 0$ and $\mathrm{ord}_{\mathfrak{q}}\, \beta \neq 0$, then $\mathfrak{q} \mid m$. Let p_∞ be the infinite prime of the rational field. Put $\mathfrak{a} = \prod_{\mathfrak{p} \nmid m p_\infty} \mathfrak{p}^{\mathrm{ord}_{\mathfrak{p}} \alpha}$ and $\mathfrak{b} = \prod_{\mathfrak{p} \nmid m p_\infty} \mathfrak{p}^{\mathrm{ord}_{\mathfrak{p}} \beta}$. Then

$$\left(\frac{\beta}{\mathfrak{a}}\right)_{km} \cdot \left(\frac{\alpha}{\mathfrak{b}}\right)_{km}^{-1} = \prod_{\mathfrak{p} \mid m p_\infty} (\alpha, \beta, k_{\mathfrak{p}}, m),$$

where $k_{\mathfrak{p}}$ is the $\mathfrak{p}$-adic completion of k.

VII. If $m > 1$ then

$$\prod_{\mathfrak{p} \mid p_\infty} (\alpha, \beta, k_{\mathfrak{p}}, m) = (-1)^{N(k, \alpha, \beta)},$$

where $N(k, \alpha, \beta)$ is the number of isomorphisms σ of k into the field of real numbers such that $\sigma\alpha < 0$, $\sigma\beta < 0$. If $m > 2$ then $N(k, \alpha, \beta) = 0$ since k contains primitive m-th roots of unity and must therefore be a complex field.

An immediate consequence of VI and VII is:

VIII. If α, β, and m are relatively prime in pairs and if $m > 1$, then

$$\left(\frac{\beta}{\alpha}\right)_{km} \cdot \left(\frac{\alpha}{\beta}\right)_{km}^{-1} = (-1)^{N(k, \alpha, \beta)} \prod_{\mathfrak{p} \mid m} (\alpha, \beta, k_{\mathfrak{p}}, m).$$

IX. If $\bar{K}$ is a finite extension of $\bar{k}$, $\beta \in \bar{k}$, then

$$(\alpha, \beta, \bar{K}, m) = (N_{\bar{K}/\bar{k}}\alpha, \beta, \bar{k}, m).$$

IX follows from the known result $\left(\frac{K\Omega/K}{\mathfrak{a}}\right) = \left(\frac{\Omega/k}{N_{\bar{K}/\bar{k}}\mathfrak{a}}\right)$. A special case of IX is:

X. If $\bar{k}$ does not contain i, $\bar{K} = \bar{k}(i)$, and if α, β, and γ are non-zero elements of $\bar{k}$, then $(\alpha^2 + \beta^2, \gamma, \bar{k}, 2) = (\alpha - \beta i, \gamma, \bar{K}, 2)$.

The problem of evaluating the reciprocity quotient is reduced by VIII to the problem of evaluating the norm residue symbol $(\alpha, \beta, k_{\mathfrak{p}}, m)$, where $\mathfrak{p} \mid m$. It is sufficient to consider the case $m = l^n$, l prime, $n \geqq 1$. For this case we need the following theorem:

THEOREM 1. *Let $\bar{R}$ be the completion of the rational field R at a finite prime l. Let $\bar{k}$ be a finite extension of $\bar{R}$ containing the l^n-th roots of unity* *exactly,*[6] *where* $l^n > 2$. *Put* $\bar{K} = \bar{k}(\zeta)$, ζ *a primitive* l^{n+s} *root of unity,* $s \geqq 1$. *Then*

$$(\alpha, \beta, \bar{k}, l^n) = (\alpha, \beta, \bar{K}, l^{n+s})^E,$$

where $E = 1$ *if* $l > 2$, *and* $E = 1 + 2^{n-1}$ *if* $l = 2$.

Proof. Let $\mathfrak{p}$ and $\mathfrak{P}$ be the prime ideals of $\bar{k}$ and $\bar{K}$ respectively. Let K' be an algebraic number field with a prime $\mathfrak{P}'$ such that $\bar{K}$ is the $\mathfrak{P}'$-adic completion of K'. Put $K = K'(\zeta)$. Then K has a prime $\mathfrak{P}$ such that $\bar{K}$ is the $\mathfrak{P}$-adic completion of K. Put $k = K \cap \bar{k}$. Then k contains the l^n-th roots of unity exactly. Furthermore k contains a prime ideal $\mathfrak{p}$ such that $\bar{k}$ is the $\mathfrak{p}$-adic completion of k, K/k is isomorphic to $\bar{K}/\bar{k}$, and $K = k(\zeta)$. By the approximation theorem α' and β' can be chosen in k such that:

(1) α'/α and β'/β are l^{n+s} powers in $\bar{k}$ and therefore in $\bar{K}$,

(2) if $\mathfrak{q} \neq \mathfrak{p}$, then $\operatorname{ord}_{\mathfrak{q}} \alpha' \neq 0$ or $\operatorname{ord}_{\mathfrak{q}} \beta' \neq 0$, and

(3) α' is an l^{n+s} power in $k_{\mathfrak{q}}$ for all primes $\mathfrak{q} \neq \mathfrak{p}$, $\mathfrak{q} \mid lp_\infty$.

From (1) it follows that $(\alpha, \beta, \bar{k}, l^n) = (\alpha', \beta', \bar{k}, l^n) = (\alpha', \beta', k_{\mathfrak{p}}, l^n)$. From (3) it follows that $(\alpha', \beta', k_{\mathfrak{q}}, l^n) = 1$ if $\mathfrak{q} \neq \mathfrak{p}$, $\mathfrak{q} \mid lp_\infty$. Put $\mathfrak{a} = \prod_{\mathfrak{q} \nmid lp_\infty} \mathfrak{q}^{\operatorname{ord}_{\mathfrak{q}} \alpha'}$ and $\mathfrak{b} = \prod_{\mathfrak{q} \nmid lp_\infty} \mathfrak{q}^{\operatorname{ord}_{\mathfrak{q}} \beta'}$. Then by VI we have

$$(\alpha, \beta, \bar{k}, l^n) = \prod_{\mathfrak{q} \mid lp_\infty} (\alpha', \beta', k_{\mathfrak{q}}, l^n) = \left(\frac{\beta'}{\mathfrak{a}}\right)_{k l^n} \cdot \left(\frac{\alpha'}{\mathfrak{b}}\right)_{k l^n}^{-1}.$$

Similarly

$$(\alpha, \beta, \bar{K}, l^{n+s}) = \left(\frac{\beta'}{\mathfrak{a}}\right)_{K l^{n+s}} \cdot \left(\frac{\alpha'}{\mathfrak{b}}\right)_{K l^{n+s}}^{-1}.$$

In the preceding paper it was proved that if k is an algebraic number field containing the l^n-th roots of unity exactly and $K = k(\zeta)$, then $\left(\frac{\gamma}{\mathrm{A}}\right)_{K l^{n+s}} = \left(\frac{\gamma}{\mathrm{A}}\right)_{k l^n}^E$, where $E = 1$ if $l > 2$, and $E = 1 + 2^{n-1}$ if $l = 2$, $n > 1$. Applying this result we obtain $(\alpha, \beta, \bar{K}, l^{n+s}) = (\alpha, \beta, \bar{k}, l^n)^E$. Now $E^2 \equiv 1 \pmod{l^n}$. Therefore $(\alpha, \beta, \bar{k}, l^n) = (\alpha, \beta, \bar{k}, l^n)^{E^2} = (\alpha, \beta, \bar{K}, l^{n+s})^E$.

Simple expressions for (α, β) are known in a number of special cases.[7] In particular if ω is a root of unity of order prime to m then clearly

$$(4) \qquad (\alpha, \omega, \bar{k}, m) = 1.$$

If ϵ is a root of unity whose order is a power of the prime l, $\mathfrak{p}$ the prime of $\bar{k}$, $\mathfrak{p} \mid l$, $l^n > 2$, and if $\alpha \equiv 1 \pmod{\mathfrak{p}}$ is an element of $\bar{k}$, then

---

[6] When a field k contains primitive l^n-th roots of unity but not primitive l^{n+1}-st roots of unity, k is said to contain the l^n-th roots of unity exactly. Cf. the preceding paper.

[7] See [4], part II, pp. 74-113, and [1].

$$(5) \qquad (\alpha, \epsilon, \bar{k}, l^n) = \epsilon^{E l^{-n} S_{\bar{k}/\bar{R}} \log \alpha},$$

where $\bar{R}$ is the l-adic completion of the rational field, $E = 1$ if $l > 2$, $E = 1 + 2^{n-1}$ if $l = 2$, and $\log \alpha$ is the $\mathfrak{p}$-adic logarithm defined by

$$\log \alpha = -\sum_{\nu=1}^{\infty} (1 - \alpha)^{\nu}/\nu.$$

If $\bar{L} = \bar{R}(\zeta)$, ζ a primitive l^n-th root of unity, $\Theta = 1/(\zeta - 1)$, and $1 - \alpha$ belongs to the prime ideal of $\bar{L}$, then

$$(6) \qquad (\alpha, 1 - \zeta, \bar{L}, l^n) = \zeta^{l^{-n} S_{\bar{L}/\bar{R}}(\zeta \Theta \log \alpha)}.$$

A weak consequence of (5) and (6) is:

LEMMA 1. *Let $\bar{L} = \bar{R}(\zeta)$, where ζ is a primitive l^n-th root of unity. Let $\bar{k}$ be a finite extension of $\bar{L}$ with the prime ideal $\mathfrak{p}$. Then if $\gamma \in \mathfrak{p}$,*

$$l^n \mid S_{\bar{k}/\bar{R}}\{\sum_{\nu=0}^{\infty} \gamma^{l^\nu} l^{-\nu}/(\zeta - 1)\}.$$

Proof. Using well known properties of the $\mathfrak{p}$-adic logarithm[8] we obtain

$$\begin{aligned} S_{\bar{k}/\bar{R}}\{\sum_{\nu=0}^{\infty} \gamma^{l^\nu} l^{-\nu}/(\zeta - 1)\} &= S_{\bar{L}/\bar{R}}(\Theta S_{\bar{k}/\bar{L}} \sum_{\nu=0}^{\infty} \gamma^{l^\nu} l^{-\nu}) \\ &= - S_{\bar{L}/\bar{R}}(\Theta S_{\bar{k}/\bar{L}} \log \beta) = S_{\bar{L}/\bar{R}}((1 - \zeta\Theta) \log N_{\bar{k}/\bar{L}} \beta) \\ &= S_{\bar{L}/\bar{R}} \log \delta - S_{\bar{L}/\bar{R}}(\zeta \Theta \log \delta), \end{aligned}$$

for suitable $\beta \in \bar{k}$, $\delta \in \bar{L}$. The terms of this last expression are divisible by l^n by (5) and (6) respectively.

3. Computation of $(1 - \omega\pi^{al^r}, \pi)$. Let l^n be any prime power greater than 2, $\bar{R}$ the completion of the rational field at the prime l, and $\bar{k}$ any finite extension of $\bar{R}$ containing the l^n-th roots of unity exactly. Let π be any local uniformizing parameter of $\bar{k}$, and ω any root of unity contained in $\bar{k}$, whose order m is relatively prime to l. Let a be a positive integer not divisible by l. The expression $(1 - \omega\pi^{al^r}, \pi, \bar{k}, l^n)$ will be computed.

Let ω_0 be the m-th root of unity satisfying $\omega_0{}^{l^r} = \omega$. We write $\omega_0 = \omega^{l^{-r}}$. Let ζ be a primitive l^n-th root of unity. If $n \geqq r$ put $\epsilon = \zeta^{l^{n-r}}$, and if $n < r$ let ϵ be a primitive l^r-th root of unity satisfying $\zeta = \epsilon^{l^{r-n}}$. In either case $\epsilon^{\xi l^{r-n}}$ shall mean ζ^{ξ}, for integral ξ. Put $\bar{K} = \bar{k}(\epsilon)$ and $n' = \operatorname{Max}(n, r)$. Then we have $(\bar{K} : \bar{k}) = l^{n'-n}$. We will write S and S' for $S_{\bar{k}/\bar{R}}$ and $S_{\bar{K}/\bar{R}}$ respectively.

Put $E' = \begin{cases} 1 + 2^{n-1} & \text{if } l = 2, r > n, \\ 1 & \text{otherwise,} \end{cases}$ and $E'' = \begin{cases} 1 + 2^{n'-1} & \text{if } l = 2, \\ 1 & \text{if } l > 2. \end{cases}$

---

[8] See [5].

Then by Theorem 1, properties I, II, and III of the norm residue symbol, and (4) and (5) we have

$$(1-\omega\pi^{al^r}, \pi, \bar{k}, l^n)^a = (1-\omega\pi^{al^r}, \pi, \bar{K}, l^{n'})^{aE'} = \prod_{\nu=0}^{l^r-1} (1-\epsilon^\nu\omega_0\pi^a, \pi^a, \bar{K}, l^{n'})^{E'}$$

$$= \prod_{\nu=0}^{l^r-1} (1-\epsilon^\nu\omega_0\pi^a, \epsilon^{-\nu}\omega_0{}^{-1}, \bar{K}, l^{n'})^{E'} = \prod_{\nu=0}^{l^r-1} (1-\epsilon^\nu\omega_0\pi^a, \epsilon, \bar{K}, l^{n'})^{-\nu E'}$$

$$= \prod_{\nu=0}^{l^r-1} \epsilon^{-\nu E'E'' l^{-n'} S' \log(1-\epsilon^\nu\omega_0\pi^a)}.$$

Now if $l > 2$ then $E'E'' = 1 = E$, and if $l = 2$ then $E'E'' \equiv 1 + 2^{n-1} = E \pmod{l^n}$. Thus in either case $E'E'' \equiv E \pmod{l^n}$. Now $(1-\omega\pi^{al^r}, \pi, \bar{k}, l^n)$ is an integral power of ζ, say ζ^A. Then modulo l^n we have

$$aA \equiv -El^{n-r-n'} \sum_{\nu=0}^{l^r-1} \nu S' \log(1-\epsilon^\nu\omega_0\pi^a)$$

$$= El^{n-r-n'} S' \{ \sum_{\nu=0}^{l^r-1} \sum_{\mu=1}^{\infty} \nu\epsilon^{\nu\mu}\omega_0{}^\mu\pi^{a\mu}/\mu \} = ES\{ \sum_{\mu=1}^{\infty} \omega_0{}^\mu\pi^{a\mu}F_1(\mu)/\mu \}$$

where $F_1(\mu) = l^{n-n'-r} S_{\bar{K}/\bar{k}} \{ \sum_{\nu=0}^{l^r-1} \nu\epsilon^{\nu\mu} \}$. If $\epsilon^{\nu\mu} \varepsilon \bar{k}$ then $S_{\bar{K}/\bar{k}}\epsilon^{\nu\mu} = (\bar{K}:\bar{k})\epsilon^{\nu\mu} = l^{n'-n}\epsilon^{\nu\mu}$, and if $\epsilon^{\nu\mu} \not\varepsilon \bar{k}$ then $S_{\bar{K}/\bar{k}}\epsilon^{\nu\mu} = 0$. Therefore $F_1(\mu) = l^{-r} \sum \nu\epsilon^{\nu\mu}$, where the summation is over those ν for which $\epsilon^{\nu\mu} \varepsilon \bar{k}$, $0 \leqq \nu < l^r$.

Now $\sum_{\xi=0}^{l^s-1} x^\xi = (x^{l^s}-1)/(x-1)$. Differentiating and multiplying by x we obtain

$$\sum_{\xi=0}^{l^s-1} \xi x^\xi = l^s x^{l^s}/(x-1) - x(x^{l^s}-1)/(x-1)^2.$$

Therefore if $\lambda^{l^s} = 1$, $\lambda \neq 1$, then

$$\sum_{\xi=0}^{l^s-1} \xi\lambda^\xi = l^s/(\lambda-1). \tag{7}$$

i) Suppose $l^r \mid \mu$. Then $\epsilon^{\nu\mu} = 1 \varepsilon \bar{k}$ for all ν, and

$$F_1(\mu) = l^{-r} \sum_{\nu=0}^{l^r-1} \nu = \tfrac{1}{2}(l^r-1).$$

ii) Suppose $l^r \nmid \mu$, $l^{n'-n} \mid \mu$. Then $\epsilon^{\nu\mu} \varepsilon \bar{k}$ for all ν, and by (7)

$$F_1(\mu) = l^{-r} \sum_{\nu=0}^{l^r-1} \nu\epsilon^{\nu\mu} = 1/(\epsilon^\mu - 1).$$

iii) Suppose $\mu = l^{n'-n-b}\mu_0$, $b > 0$, $l \nmid \mu_0$. Then $\epsilon^{\nu\mu} \in \bar{k}$ if and only if $l^b \mid \nu$. Making the substitution $\nu = l^b\xi$, and using (7)

$$F_1(\mu) = l^{-r} \sum \nu\epsilon^{\nu\mu} = l^{-r} \sum_{\xi=0}^{l^{r-b}-1} l^b\xi\epsilon^{l^b\xi\mu} = 1/(\epsilon^{\mu l^b} - 1).$$

$$\text{Summarizing, } F_1(\mu) = \begin{cases} \tfrac{1}{2}(l^r - 1) & \text{if } l^r \mid \mu, \\ 1/\epsilon^{\mu} - 1) & \text{if } l^r \nmid \mu,\ l^{n'-n} \mid \mu, \\ 1/\epsilon^{\mu l^b} - 1) & \text{if } l^{n'-n-b} \parallel \mu,\ b > 0. \end{cases}$$

Define $F_2(\mu) = -\tfrac{1}{2}$ if $l^r \mid \mu$, and $F_2(\mu) = F_1(\mu)$ if $l^r \nmid \mu$. Then

$$\sum_{\mu=1}^{\infty} \omega_0{}^{\mu}\pi^{a\mu}[F_1(\mu) - F_2(\mu)]/\mu = \sum_{l^r \mid \mu} \tfrac{1}{2}\omega_0{}^{\mu}\pi^{a\mu}l^r/\mu = -\tfrac{1}{2}\log(1 - \omega_0{}^{l^r}\pi^{al^r})$$
$$= -\tfrac{1}{2}\log(1 - \omega\pi^{al^r}).$$

Therefore $A \equiv ES\{\sum_{\mu=1}^{\infty} \omega_0{}^{\mu}\pi^{a\mu}F_2(\mu)/(a\mu)\} - \tfrac{1}{2}(E/a)\, S\log(1 - \omega\pi^{al^r}) \pmod{l^n}$.

Now by (5), if $\alpha \equiv 1 \pmod{\pi}$, then $l^n \mid ES\log\alpha$. Since $l \nmid E$, $l \nmid a$, it follows that $-\tfrac{1}{2}(E/a)S\log\alpha \equiv \tfrac{1}{2}S\log\alpha \pmod{l^n}$ for all $\alpha \equiv 1 \pmod{\pi}$. Therefore $A \equiv ES\{\sum_{\mu=1}^{\infty} \omega_0{}^{\mu}\pi^{a\mu}F_2(\mu)/(a\mu)\} + \tfrac{1}{2}S\log(1 - \omega\pi^{al^r}) \pmod{l^n}$. Furthermore if $l > 2$ the second term can be dropped. Now we put $\Delta(\mu) = 1/(\zeta^c - 1)$, where $\mu = l^{\nu}c$, $l \nmid c$. Then by Lemma 1

$$S\{\sum_{\mu=1}^{\infty} \omega_0{}^{\mu}\pi^{a\mu}\Delta(\mu)/(a\mu)\} = \sum_{l \nmid c}(ac)^{-1}S\{\sum_{\nu=0}^{\infty}(\omega_0{}^{c}\pi^{ac})^{l^{\nu}}l^{-\nu}/(\zeta^c - 1)\} \equiv 0 \pmod{l^n}.$$

We note that if $l^{n'-n} \nmid \mu$ then $F_2(\mu) = \Delta(\mu)$. We now adopt the convention that $S(\gamma\pi^t) = 0$ if γ is a non-zero element of $\bar{k}$ and t is not an integer. Then making the substitution $\mu = \xi l^{r-n}$ we obtain, modulo l^n,

$$ES\{\sum_{\mu=1}^{\infty} \omega_0{}^{\mu}\pi^{a\mu}F_2(\mu)/(a\mu)\} \equiv ES\{\sum_{\mu=1}^{\infty} \omega_0{}^{\mu}\pi^{a\mu}[F_2(\mu) - \Delta(\mu)]/(a\mu)\}$$
$$= ES\{\sum_{l^{n'-n} \mid \mu} \omega_0{}^{\mu}\pi^{a\mu}[F_2(\mu) - \Delta(\mu)]/(a\mu)\}$$
$$= ES\{\sum_{\xi=1}^{\infty} \omega_0{}^{\xi l^{r-n}}\pi^{a\xi l^{r-n}}[F_2(\xi l^{r-n}) - \Delta(\xi l^{r-n})]/(a\xi l^{r-n})\}$$
$$= ES\{\sum_{\xi=1}^{\infty} \omega^{\xi l^{-n}}\pi^{a\xi l^{r-n}}F_{\zeta}(\xi)/(a\xi l^{r-n})\},$$

where $F_{\zeta}(\xi) = F_2(\xi l^{r-n}) - \Delta(\xi l^{r-n})$. If $\xi = cl^{v}$, $l \nmid c$, then

$$F_{\zeta}(\xi) = \begin{cases} 1/(\zeta^{\xi} - 1) - 1/(\zeta^c - 1) & \text{if } l^n \nmid \xi. \\ -\tfrac{1}{2} - 1/(\zeta^c - 1) & \text{if } l^n \mid \xi, \end{cases}$$

Combining results we have

$$(8)\qquad \begin{aligned} &(1-\omega\pi^{al^r}, \pi, \bar{k}, l^n) = \zeta^A, \\ &A = ES\{\sum_{\xi=1}^{\infty} \omega^{\xi l^{-n}}\pi^{a\xi l^{r-n}}F_\xi(\xi)/(a\xi l^{r-n}) + \tfrac{1}{2}S\log(1-\omega\pi^{al^r})\}. \end{aligned}$$

In deriving (8) the assumption was made that $l^n > 2$. A similar result will now be derived for the case $l^n = 2$.

Let $\bar{R}$ be the completion of the rational field at the prime 2. Let $\bar{k}$ be a finite extension of $\bar{R}$ not containing i. Let π be a local uniformizing parameter of $\bar{k}$, and ω be a root of unity of odd order. We shall begin by computing $(1+\omega\pi^{a2^r}, \pi, \bar{k}, 2)$, where a is odd and $r \geqq 1$. Put $\bar{L} = \bar{k}(i)$. Let $\bar{L}$ contain the $2^{2+\delta}$ roots of unity exactly, and ρ be a primitive $2^{2+\delta}$ root of unity satisfying $\rho^{2^\delta} = i$. If $r \leqq \delta + 1$ put $\epsilon = \rho^{2^{\delta+1-r}}$, and if $r > \delta + 1$ let ϵ be a primitive 2^{r+1}-st root of unity satisfying $\epsilon^{2^{r-1-\delta}} = \rho$. In either case $\epsilon^{\xi 2^{r-1-\delta}}$ shall mean ρ^ξ, for integral ξ. Put $\bar{K} = \bar{L}(\epsilon)$. We will write S, S', and $\bar{S}$ for $S_{\bar{k}/\bar{R}}$, $S_{\bar{K}/\bar{R}}$, and $S_{\bar{L}/\bar{k}}$ respectively. Put $\omega_0 = \omega^{2^{-r}}$, that is ω_0 is the root of unity of odd order satisfying $\omega_0^{2^r} = \omega$. Put $n' = \mathrm{Max}(r+1, 2+\delta)$. Then by properties V and X the norm residue symbol and Theorem 1 we have

$$\begin{aligned} (1+\omega\pi^{a2^r}, \pi, \bar{k}, 2) &= (1-i\omega_0^{2^{r-1}}\pi^{a2^{r-1}}, \pi, \bar{L}, 2) \\ &= (1-i\omega_0^{2^{r-1}}\pi^{a2^{r-1}}, \pi, \bar{L}, 2^{2+\delta})^{2^{1+\delta}} = (1-i\omega_0^{2^{r-1}}\pi^{a2^{r-1}}, \pi, \bar{K}, 2^{n'})^{2^{1+\delta}E'}, \end{aligned}$$

where $E' \equiv 1 \pmod 2$. Since $(1+\omega\pi^{a2^r}, \pi, \bar{k}, 2)$ is a square root of unity E' may be replaced by 1. Proceeding as in the case $l^n > 2$,

$$\begin{aligned} (1+\omega\pi^{a2^r}, \pi, \bar{k}, 2) &= (1-i\omega_0^{2^{r-1}}\pi^{a2^{r-1}}, \pi, \bar{K}, 2^{n'})^{2^{1+\delta}} \\ &= \prod_{\nu=0}^{2^{r-1}-1} (1-\epsilon^{4\nu+1}\omega_0\pi^a, \pi, \bar{K}, 2^{n'})^{2^{1+\delta}} \\ &= \prod_{\nu=0}^{2^{r-1}-1} (1-\epsilon^{4\nu+1}\omega_0\pi^a, \epsilon, \bar{K}, 2^{n'})^{-2^{1+\delta}}(4\nu+1)/a = (-1)^{\mathrm{A}}, \end{aligned}$$

$$\begin{aligned} aA &\equiv -(1+2^{n'-1})2^{-r-n'+1+\delta}\sum_{\nu=0}^{2^{r-1}-1}(4\nu+1)S'\log(1-\epsilon^{4\nu+1}\omega_0\pi^a) \\ &\equiv 2^{-r-n'+1+\delta}\sum_{\nu=0}^{2^{r-1}-1}(4\nu+1)S'\{\sum_{\mu=1}^{\infty}\epsilon^{(4\nu+1)\mu}\omega_0^{\mu}\pi^{a\mu}/\mu\} \\ &= S\{\sum_{\mu=1}^{\infty}\omega_0^{\mu}\pi^{a\mu}F_1(\mu)/\mu\} \qquad \pmod 2, \end{aligned}$$

where $F_1(\mu) = 2^{-r-n'+1+\delta}\bar{S}\{\sum_{\nu=0}^{2^{r-1}-1}(4\nu+1)S_{\bar{K}/\bar{L}}\epsilon^{(4\nu+1)\mu}\}$.

i) Suppose $2^{r-1} \mid \mu$. Then $\epsilon^{\mu} \varepsilon \bar{L}$ and $\epsilon^{4\nu\mu} = 1$ for all [illegible]. Therefore $S_{\bar{K}/\bar{L}}\epsilon^{(4\nu+1)\mu} = (\bar{K}:\bar{L})\epsilon^{\mu} = 2^{n'-2-\delta}\epsilon^{\mu}$. Thus in this case

$$F_1(\mu) = 2^{-r-1} \sum_{\nu=0}^{2^{r-1}-1} (4\nu + 1)\bar{S}\epsilon^{\mu} = \tfrac{1}{4}(2^r - 1)\bar{S}\epsilon^{\mu}$$

$$= \begin{cases} \tfrac{1}{2}(-1)^{\mu 2^{-r}}(2^r - 1) & \text{if } 2^r \mid \mu, \\ 0 & \text{if } 2^{r-1} \parallel \mu. \end{cases}$$

ii) Suppose $2^{r-1} \nmid \mu$, $2^{r-1-\delta} \mid \mu$. Then $\epsilon^{(4\nu+1)\mu} \varepsilon \bar{L}$, $\epsilon^{4\mu} \neq 1$. Using (7) we have

$$F_1(\mu) = 2^{-r-1}\bar{S}(\epsilon^{\mu} \sum_{\nu=0}^{2^{r-1}-1} (4\nu + 1)\epsilon^{4\nu\mu})$$

$$= 2^{-r-1}\bar{S}(\epsilon^{\mu} \sum_{\nu=0}^{2^{r-1}-1} 4\nu\epsilon^{4\nu\mu}) = \bar{S}(\epsilon^{\mu}/(\epsilon^{4\mu} - 1)).$$

iii) Suppose $2^{r-1-\delta} \nmid \mu$. Then $\epsilon^{\mu} \not\in \bar{L}$ and so $\epsilon^{(4\nu+1)\mu} \not\in \bar{L}$ which implies that $\bar{S}\epsilon^{(4\nu+1)\mu} = 0$. Therefore $F_1(\mu) = 0$. Summarizing

$$(9) \qquad F_1(\mu) = \begin{cases} \tfrac{1}{2}(-1)^{\mu 2^{-r}}(2^r - 1 & \text{if } 2^r \mid \mu, \\ \bar{S}(\epsilon^{\mu}/(\epsilon^{4\mu} - 1)) & \text{if } 2^{r-1} \nmid \mu,\ 2^{r-1-\delta} \mid \mu, \\ 0 & \text{otherwise.} \end{cases}$$

We now make the assertion that for $2 \nmid a$, $r \geqq 0$,

$$(10) \qquad (1 + \omega(-\pi)^{a2^r}, \pi, \bar{k}, 2) = (-1)^{A},$$

where $A \equiv S\{\sum_{\mu=1}^{\infty} \omega_0{}^{\mu}\pi^{a\mu}F_1(\mu)/(a\mu)\} \pmod 2$, $F_1(\mu)$ defined by (9). If $r \geqq 1$, then $(-\pi)^{a2^r} = \pi^{a2^r}$ and (10) follows immediately. If $r = 0$, then

$$(1 + \omega(-\pi)^{a2^r}, \pi, \bar{k}, 2) = (1 - \omega\pi^{a}, \pi, \bar{k}, 2) = (1 - \omega\pi^{a}, \omega, \bar{k}, 2)^{-1/a} = 1.$$

In this case $2^r = 1$, and hence $F_1(\mu) = 0$ for all μ. Thus A is even and the proof of (10) is complete. Put $F_2(\mu) = -\tfrac{1}{2}(-1)^{\mu 2^{-r}}$ if $2^r \mid \mu$, and $F_2(\mu) = F_1(\mu)$ if $2^r \nmid \mu$. Then

$$S\{\sum_{\mu=1}^{\infty} \omega_0{}^{\mu}\pi^{a\mu}[F_1(\mu) - F_2(\mu)]/(a\mu)\} = S\{\sum_{2^r \mid \mu} \omega_0{}^{\mu}\pi^{a\mu}(-1)^{\mu 2^{-r}}2^{r-1}/(a\mu)\}$$

$$\equiv S\{\sum_{\xi=1}^{\infty} \tfrac{1}{2}(-1)^{\xi}\omega_0{}^{\xi 2^r}\pi^{a\xi 2^r}/\xi\} = -\tfrac{1}{2}S\log(1 + \omega_0{}^{2^r}\pi^{a2^r})$$

$$= -\tfrac{1}{2}\log N_{\bar{k}/\bar{R}}(1 + \omega_0{}^{2^r}\pi^{a2^r}) \equiv 0 \qquad \pmod 2,$$

since 4 divides the 2-adic logarithm of any odd integer in $\bar{R}$. As before we adopt the convention that $S(\gamma\pi^{t}) = 0$ if γ is a non-zero element of $\bar{k}$ and t is not an integer. Then we have, modulo 2,

$$\begin{aligned} A &\equiv S\{ \sum_{\mu=1}^{\infty} \omega_0{}^{\mu}\pi^{a\mu}F_2(\mu)/(a\mu) \} \\ &= S\{ \sum_{\xi=1}^{\infty} \omega_0{}^{\xi 2^{r-1-\delta}}\pi^{a\xi 2^{r-1-\delta}}F_2(\xi 2^{r-1-\delta})/(a\xi 2^{r-1-\delta}) \} \\ &= S\{ \sum_{\xi=1}^{\infty} \omega^{\xi 2^{-1-\delta}}\pi^{a\xi 2^{r-1-\delta}}F_{-1}(\xi)/(a\xi 2^{r-1-\delta}) \}, \end{aligned}$$

where

$$F_{-1}(\xi) = \begin{cases} -\tfrac{1}{2}(-1)^{\xi 2^{-1-\delta}} & \text{if } 2^{1+\delta} \mid \xi, \\ 0 & \text{if } 2^{\delta} \parallel \xi, \\ \bar{S}(\epsilon^{\xi 2^{r-1-\delta}}/(\epsilon^{\xi 2^{r-1-\delta}} = 1)) & \text{if } 2^{\delta} \nmid \xi. \end{cases}$$

The results of this section will now be combined into one theorem.

THEOREM 2. *Let l be a rational prime and n a positive integer. Let $\bar{R}$ be the completion of the rational field at l, and let $\bar{k}$ be a finite extension of $\bar{R}$ containing the l^{n}-th roots of unity exactly. Let π be any local uniformizing parameter of $\bar{k}$, and ω any root of unity of $\bar{k}$, whose order m is relatively prime to l. Let ζ be a primitive l^{n}-th root of unity. Let the field obtained from $\bar{k}$ by adjoining the primitive l^{n+1}-st roots of unity contain the $l^{n+1+\delta}$ roots of unity exactly. If $l^{n} = 2$ let ρ be a primitive $2^{n+1+\delta}$ root of unity. If $l^{n} > 2$, then $\delta = 0$. Put*

$$\eta = \begin{cases} -1 & \text{if } l^{n} = 2, \\ +1 & \text{if } l^{n} > 2, \end{cases} \quad \textit{and } E = \begin{cases} 1 + 2^{n-1} & \text{if } l = 2, n > 1, \\ 1 & \text{otherwise.} \end{cases}$$

Let S stand for $S_{\bar{k}/\bar{R}}$, with the convention noted above, and let $\omega^{\xi l^{-t}}$ be defined as the m-th root of unity ω' satisfying $\omega'^{l^{t}} = \omega^{\xi}$. Put

$$F_{\zeta}(\xi) = \begin{cases} -\tfrac{1}{2} - 1/(\zeta^{c} - 1) & \text{if } l^{n} > 2, \xi = cl^{v}, l \nmid c, v \geqq n, \\ 1/(\zeta^{\xi} - 1) - 1/(\zeta^{c} - 1) & \text{if } l^{n} > 2, \xi = cl^{v}, l \nmid c, v < n, \\ -\tfrac{1}{2}(-1)^{\xi 2^{-1-\delta}} & \text{if } l^{n} = 2, 2^{1+\delta} \mid \xi, \\ 0 & \text{if } l^{n} = 2, 2^{\delta} \parallel \xi, \\ \mathrm{S}_{\bar{k}(i)/\bar{k}}(\rho^{\xi}/(\rho^{4\xi} - 1)) & \text{if } l^{n} = 2, 2^{\delta} \nmid \xi. \end{cases}$$

Then, if $l \nmid a$,

$$(1 - \eta\omega(\eta\pi)^{al^{r}}, \pi, \bar{k}, l^{n}) = \zeta^{A},$$

where

$$A = ES\{ \sum_{\xi=1}^{\infty} \omega^{\xi l^{-n-\delta}}\pi^{a\xi l^{r-n-\delta}}F_{\zeta}(\xi)/(\mathrm{a}\xi l^{r-n-\delta}) \} + \tfrac{1}{4}(\eta + 1)S\log(1 - \eta\omega(\eta\pi)^{al^{r}}).$$

4. A formula for (α, π). We will now use Theorem 2 to derive a formula for (α, π, k, l^n), valid when $\alpha \equiv 1 \pmod{\pi}$. We need the following lemma:

LEMMA 2. *Let $\mu(N)$ be the Moebius μ function, and*

$$\mu^*(M) = \begin{cases} \mu(M) & \text{if } l^n > 2, \\ -\mu(M) & \text{if } l^n = 2, M \text{ odd}, \\ -2^{r-1}\mu(N) & \text{if } l^n = 2, M = 2^r N, N \text{ odd}, r \geqq 1. \end{cases}$$

Then

$$\sum_{d|M} \eta^{M/d}\mu^*(d) = \begin{cases} 1 \text{ if } M = 1, \\ 0 \text{ if } M > 1, \end{cases}$$

where η is as defined in Theorem 2.

Proof. Case I. $l^n > 2$. In this case $\eta = 1$, $\mu^*(d) = \mu(d)$ and therefore

$$\sum_{d|M} \eta^{M/d}\mu^*(d) = \sum_{d|M} \mu(d) = \begin{cases} 1 \text{ if } M = 1, \\ 0 \text{ if } M > 1. \end{cases}$$

Case II. $l^n = 2$. In this case $\eta = -1$. Put $M = 2^r N$, N odd. Making the substitution $d = 2^\nu D$, D odd, we obtain

$$\begin{aligned} \sum_{d|M} \eta^{M/d}\mu^*(d) &= \sum_{\nu=0}^{r} \sum_{D|N} (-1)^{2^{r-\nu}} \mu^*(2^\nu D) \\ &= -\left((-1)^{2^r} + \sum_{\nu=1}^{r} (-1)^{2^{r-\nu}} 2^{\nu-1}\right) \sum_{D|N} \mu(D). \end{aligned}$$

Now

$$(-1)^{2^r} + \sum_{\nu=1}^{r} (-1)^{2^{r-\nu}} 2^{\nu-1} = \begin{cases} -1 \text{ if } r = 0, \\ 0 \text{ if } r > 0, \end{cases}$$

and

$$\sum_{D|N} \mu(D) = \begin{cases} 1 \text{ if } N = 1, \\ 0 \text{ if } N > 1. \end{cases}$$

Lemma 2 follows immediately.

A necesary and sufficient condition for the series $\sum_{\nu=1}^{\infty} a_\nu$ to converge in k is that $\operatorname{ord}_{\mathfrak{p}} a_\nu \to \infty$ as $\nu \to \infty$, where $\mathfrak{p}$ is the prime of k. Furthermore any convergent series in k is absolutely convergent. It follows that we can manipulate with much greater freedom than in ordinary analysis. In particular

$$
\begin{aligned}
&E \sum_{\xi=1}^{\infty} \omega^{\xi l^{-n-\delta}} \pi^{a\xi l^{r-n-\delta}} F_\zeta(\xi)/(a\xi l^{r-n-\delta}) \\
&\quad = E \sum_{\xi=1}^{\infty} F_\zeta(\xi) \sum_{\phi=1}^{\infty} (\omega^{\phi\xi l^{-n-\delta}} \pi^{a\phi\xi l^{r-n-\delta}} / (\phi^2 a\xi l^{r-n-\delta})) \sum_{\nu|\phi} \mu^*(\nu)\eta^{\phi/\nu} \\
(11) \quad &\quad = E \sum_{\xi=1}^{\infty} F_\zeta(\xi) \sum_{\nu=1}^{\infty} (\mu^*(\nu)/\nu) \sum_{\theta=1}^{\infty} \eta^\theta \omega^{\theta\nu\xi l^{-n-\delta}} \pi^{a\theta\nu\xi l^{r-n-\delta}} / (\theta^2 \nu a\xi l^{r-n-\delta}) \\
&\quad = E \sum_{\tau=1}^{\infty} \sum_{\nu|\tau} (F_\zeta(\tau/\nu)\mu^*(\nu)/\nu) \sum_{\theta=1}^{\infty} \eta^\theta \omega^{\theta\tau l^{-n-\delta}} \pi^{a\theta\tau l^{r-n-\delta}} / a\theta^2\tau l^{r-n-\delta}) \\
&\quad = \sum_{\tau=1}^{\infty} J_\zeta(\tau) \sum_{\theta=1}^{\infty} \eta^\theta \omega^{\theta\tau l^{-n-\delta}} \pi^{a\theta\tau l^{r-n-\delta}} / (a\theta^2\tau l^{r-n-\delta}),
\end{aligned}
$$

where $J_\zeta(\tau) = E \sum_{\nu|\tau} F_\zeta(\tau/\nu)\mu^*(\nu)/\nu$.

Let f be the residue class field degree of the extension $\bar{k}/\bar{R}$. Let w be a primitive $l^f - 1$ root of unity. Let $\alpha \equiv 1 \pmod{\pi}$ be an element of $\bar{k}$. Let $G(x, y) = 1 + \sum_{\nu=1}^{\infty} \sum_{\mu=0}^{l^f-2} a_{\nu\mu} x^\mu y^\nu$, where the $a_{\nu\mu}$ are rational integers such that $G(w, \eta\pi) = \alpha$. Then $\log G(x, y) = -\sum_{\nu=1}^{\infty} [1 - G(x, y)]^\nu/\nu$ is a power series in x and y with rational coefficients, say

$$\log G(x, y) = \sum_{\nu=1}^{\infty} \sum_{\mu=0}^{\nu(l^f-2)} b_{\nu\mu} x^\mu y^\nu.$$

Definition. $[\alpha, \tau l^{-t} \mid G, \pi] = -\sum_{v=1}^{\infty} \sum_{\mu=0}^{v(l^f-2)} b_{\nu\mu} w^{\mu\tau l^{-t}} \pi^{\nu\tau l^{-t}} / (\nu\tau l^{-t})$.

It is easily seen that the series for $[\alpha, \tau l^{-t} \mid G, \pi]$ converges and that $[\alpha_1, \tau l^{-t} \mid G_1, \pi] + [\alpha_2, \tau l^{-t} \mid G_2, \pi] = [\alpha_1\alpha_2, \tau l^{-t} \mid G_1 G_2, \pi]$.

THEOREM 3. *Let* $G(x, y) = 1 + \sum_{\nu=1}^{\infty} \sum_{\mu=0}^{l^f-2} a_{\nu\mu} x^\mu y^\nu$, $a_{\nu\mu}$ *rational integers such that* $G(w, \eta\pi) = \alpha$. *Then*

$$(12) \qquad (\alpha, \pi, \bar{k}, l^n) = \zeta^B,$$

where

$$B = S\{ \sum_{\tau=1}^{\infty} J_\zeta(\tau) [\alpha, \tau l^{-n-\delta} \mid G, \pi]\} + \tfrac{1}{4}(\eta + 1) S \log \alpha.$$

Proof. $G(x, y)$ can be written as an infinite product

$$G(x, y) = \prod_{v=1}^{\infty} \prod_{u=0}^{v(l^f-2)} \{G_{vu}(x, y)^{c_{vu}},$$

where $G_{vu}(x, y) = 1 - \eta x^u y^v$, and the c_{vu} are rational integers. Since both sides of (12) are multiplicative in α, it is sufficient to prove (12) for the case $G(x, y) = 1 - \eta x^u y^v$. In this case the desired result follows immediately from (11) and Theorem 2.

5. The reciprocity quotient. Let m be a rational integer greater than 1, and let k be an algebraic number field containing primitive m-th roots of unity. Let $m = \prod l_\nu^{n_\nu}$ be the factorization of m into rational primes. Let a_ν be rational integers such that $\sum a_\nu m l_\nu^{-n_\nu} = 1$. Let $N(k, \alpha, \beta)$ be the number of isomorphisms σ of k into the field of real numbers such that $\sigma\alpha < 0$, $\sigma\beta < 0$. For each prime ideal $\mathfrak{p}$ dividing m let $k_\mathfrak{p}$ be the $\mathfrak{p}$-adic completion of k. If $\mathfrak{p}$ divides the rational prime l, let $k_\mathfrak{p}$ contain the $l^{n_\mathfrak{p}}$-th roots of unity exactly, and let $\zeta_\mathfrak{p}$ be a primitive $l^{n_\mathfrak{p}}$-th root of unity. If α and β are elements of k that are relatively prime to each other and to m, we have

$$\left(\frac{\alpha}{\beta}\right)_{km} \cdot \left(\frac{\beta}{\alpha}\right)_{km}^{-1} = \prod \left\{ \left(\frac{\alpha}{\beta}\right)_{kl_\nu^{n_\nu}} \cdot \left(\frac{\beta}{\alpha}\right)_{kl_\nu^{n_\nu}}^{-1} \right\}^{a_\nu}$$
$$= (-1)^{N(k,\alpha,\beta)} \prod_{\nu} \prod_{\mathfrak{p} | l_\nu} (\alpha, \beta, k_\mathfrak{p}, l_\nu^{n_\nu})^{-a_\nu}.$$

Let π be an element of k such that $\pi \varepsilon \mathfrak{p}$, $\pi \not\varepsilon \mathfrak{p}^2$ for all prime ideals $\mathfrak{p}$ dividing m. Then for all such $\mathfrak{p}$, π and $\beta\pi$ are local uniformizing parameters for $k_\mathfrak{p}$, and

$$(\alpha, \beta, k_\mathfrak{p}, l_\nu^{n_\nu}) = (\alpha, \beta\pi, k_\mathfrak{p}, l_\nu^{n_\mathfrak{p}})^{l_\nu^{n_\mathfrak{p} - n_\nu}} \cdot (\alpha, \pi, k_\mathfrak{p}, l_\nu^{n_\mathfrak{p}})^{-l_\nu^{n_\mathfrak{p} - n_\nu}}.$$

Now $\alpha^{N\mathfrak{p}-1} \equiv 1 \pmod{\mathfrak{p}}$, and if power series $G_\mathfrak{p}(x, y)$ and $H_\mathfrak{p}(x, y)$ are suitably chosen we can apply Theorem 3 and obtain:

THEOREM 4.

$$\left(\frac{\alpha}{\beta}\right)_{km} \cdot \left(\frac{\beta}{\alpha}\right)_{km}^{-1} = (-1)^{N(k,\alpha,\beta)} \prod_{\nu} \prod_{\mathfrak{p} | l_\nu} \zeta_\mathfrak{p}^{C_\mathfrak{p} a_\nu l_\nu^{n_\mathfrak{p} - n_\nu}/(N\mathfrak{p}-1)},$$

where

$$C_\mathfrak{p} = S_{k_\mathfrak{p}/R_{l_\nu}} \{ \sum_{\tau=1}^{\infty} J_{\zeta_\mathfrak{p}}(\tau) ([\alpha^{N\mathfrak{p}-1}, \tau l_\nu^{-n-\delta} \mid G_\mathfrak{p}, \pi] - [\alpha^{N\mathfrak{p}-1}, \tau l_\nu^{-n-\delta} \mid H_\mathfrak{p}, \beta\pi] \}.$$

YALE UNIVERSITY.

REFERENCES.

[1] E. Artin and H. Hasse, "Die beiden Ergänzungssätze zum Reziprozitätsgesetz der l^n-ten Potenzreste im Körper der l^n-ten Einheitswurzeln," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, vol. 6 (1928), pp. 146-162.

[2] E. Artin and G. Whaples, "Axiomatic characterization of fields by the product formula for valuations," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 469-492.

[3] C. Chevalley, "La théorie du corps de classes," *Annals of Mathematics*, vol. 41 (1940), pp. 394-418.

[4] H. Hasse, "Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper," parts I, Ia, and II, *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 35 (1926), pp. 1-55, vol. 36 (1927), pp. 233-311, and *Ergänzungsband* VI (1930) respectively.

[5] ———, "Die Gruppe der p^n-Primären Zahlen für einen Primteiler $\mathfrak{p}$ von p," *Journal für die reine und angewandte Mathematik*, vol. 176 (1937), pp. 174-183.

[6] H. Rothgiesser, "Zum Reziprozitätsgesetz für l^n," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, vol. 11 (1935-36), pp. 1-16.

[7] I. R. Šafarevič, "A general reciprocity law," *Doklady Akademii Nauk SSSR*, vol. 64 (1949), pp. 25-28 (Russian).

[8] ———, "A general reciprocity law," *Matematičeskij Sbornik*, vol. 26(68) (1950), pp. 113-146 (Russian).

EXTREMAL PROBLEMS IN THE THEORY OF BOUNDED ANALYTIC FUNCTIONS.*

By Zeev Nehari.

1. Introduction and notations. Let B be the family of analytic functions $f(z)$ which are regular and single-valued and satisfy $|f(z)| \leq 1$ in a multiply-connected domain D. Generalizing the classical case in which D is the unit sircle, it is possible to pose a large number of extremal problems for continuous functionals defined on B. Some problems of this kind were solved by H. Grunsky [5, 6], L. V. Ahlfors [1], P. R. Garabedian [4] and the author [8]. In the present paper, a method is developed which makes it possible to solve a large variety of such problems; the method works equally well when the class B is restricted by various additional conditions. In all cases, the extremal functions yield $(1, m)$ conformal maps of D onto the unit circle, where m depends on the particular extremal problem and can be explicitly stated in each case. The method consists in associating with each given problem an allied extremal problem within a family of specified functions involving a finite number of parameters. The boundary values of the solution of the associated problem are used to set up a Dirichlet problem, which is then shown to yield the solution of the original problem.

We shall use the following notations:

D denotes a finite domain in the complex z-plane bounded by n closed analytic curves $\Gamma_1, \Gamma_2, \cdots, \Gamma_n$; the boundary of D will be denoted by Γ, that is, $\Gamma = \Gamma_1 + \Gamma_2 + \cdots + \Gamma_n$. The assumption that the boundary components are analytic curves is only made for the convenience of the proofs; by standard arguments, all results can be readily extended to the case of general boundary components.

B denotes the family of analytic functions $f(z)$ which are regular, single-valued, and bounded, say $|f(z)| \leq 1$, in D.

B_R denotes the family of analytic functions $f(z)$ which are regular and single-valued and have a bounded real part, say $|\operatorname{Re}\{f(z)\}| \leq 1$, in D. The classes B and B_R can be transformed into each other by an elementary conformal mapping, and extremal problems in one class are equivalent to corresponding extremal problems in the other. Since, from the point of view

---

* Received March 13, 1950; presented to the Society, December 27, 1949.

of the method of this paper, the class B_R is much easier to handle than the class B, most proofs will be carried out for functions of B_R.

$g(z,\zeta)$ denotes the Green's function of D with respect to the point $\zeta \varepsilon D$, i. e. $g(z,\zeta)+\log|z-\zeta|$ is harmonic in D and $g(z,\zeta)$ vanishes for $z \varepsilon \Gamma$.

$N(z,\zeta)$ denotes the Neumann function of D with respect to the point $\zeta \varepsilon D$, i. e. $N(z,\zeta)+\log|z-\zeta|$ is harmonic in D and the normal derivative of $N(z,\zeta)$ is constant on Γ.

$\omega_\nu(z)$ denotes the harmonic measure of Γ_ν with respect to D, i. e. $\omega_\nu(z)$ is harmonic in D, and it has the boundary values 1 and 0 on Γ_ν and $\Gamma-\Gamma_\nu$, respectively.

2. The Schwarz lemma. In order to illustrate the method of proof in a comparatively simple case, we first consider the generalization of the classical Schwarz lemma, that is, the problem: $f(z) \varepsilon B$, $|f'(\zeta)| = \max.$ $(\zeta \varepsilon D)$ which was treated—by different methods—in [1, 4, 8]. We note that the maximizing function $F(z)$ must vanish at $z=\zeta$, since otherwise the function [1]

$$F_1(z)=\frac{F(z)-F(\zeta)}{1-[F(\zeta)]^*F(z)}=\frac{F'(\zeta)}{1-|F(\zeta)|^2}(z-\zeta)+O(|z-\zeta|^2)$$

would satisfy $|F'_1(\zeta)|>|F'(\zeta)|$. We may also assume that $F'(\zeta)>0$ since, for $|\gamma|=1$, $\gamma F(z)$ is also in B.

The transformation $w_1=(4/\pi)\tan^{-1}w$ maps the circle $|w|<1$ conformally onto the infinite strip $-1<\operatorname{Re}\{w_1\}<1$. Hence, the function

$$\phi(z)=(4/\pi)\tan^{-1}f(z) \tag{1}$$

is in B_R. Since $F(\zeta)=0$ and $F'(\zeta)>0$, we may confine ourselves to functions $f(z)$ which likewise satisfy these conditions. We then have

$$\phi(\zeta)=0, \qquad \phi'(\zeta)=(4/\pi)f'(\zeta)>0. \tag{2}$$

If $\zeta=\xi+i\eta$, the above problem is therefore equivalent to the following problem concerning the class B_R: $\phi(z) \varepsilon B_R$, $\phi(\zeta)=0$, $\operatorname{Re}\{\phi'(\zeta)\}=\max.$ We note that in carrying through the proof we may confine ourselves to functions $\phi(z)$ which are continuous in $D+\Gamma$. Once the result is obtained for this restricted class, the general case then follows by a standard argument involving the approximation of D by a sequence of domains interior to it.

---

[1] Complex conjugates are denoted by asterisks.

By Green's formula, we have

$$(3)\qquad \mathrm{Re}\{\phi(\zeta)\} = u(\zeta) = -\tfrac{1}{2\pi}\int_\Gamma u(z)\,(\partial g(z,\zeta)/\partial n)\,ds, \qquad ds = |\,dz\,|,$$

where $\phi(z) = u(z) + iv(z)$, and $\partial/\partial n$ denotes differentiation with respect to the outwards pointing normal. Differentiating (3) with respect to ξ, we obtain

$$(4)\quad \phi'(\zeta) = \mathrm{Re}\{\partial\phi(\zeta)/\partial\xi\} = \partial u(\zeta)/\partial\xi = -\tfrac{1}{2\pi}\int_\Gamma u(z)\,(\partial^2 g(z,\zeta)/\partial n\partial\xi)\,ds.$$

We also have

$$(5)\qquad \int_\Gamma u(z)\,(\partial\omega_\nu(z)/\partial n)\,ds = 0, \qquad \nu = 1,\cdots,n.$$

Indeed by Green's formula, the Cauchy-Riemann equations, and the boundary properties of $\omega_\nu(z)$,

$$(5')\qquad \int_\Gamma u(\partial\omega_\nu/\partial n)\,ds = \int_\Gamma \omega_\nu(\partial u/\partial n)\,ds = \int_{\Gamma_\nu}(\partial u/\partial n)\,ds$$
$$= \int_{\Gamma_\nu}(\partial v/\partial s)\,ds = \int_{\Gamma_\nu} dv,$$

and this vanishes, since the function $\phi(z) = u(z) + iv(z)$ is single-valued in D. We further note that

$$(6)\qquad \int_\Gamma u(z)\,(\partial g(z,\zeta)/\partial n)\,ds = 0,$$

since $u(\zeta) = 0$. It follows from (4), (5), and (6) that

$$(7)\quad u_\xi(\zeta) = \phi'(\zeta)$$
$$= -\tfrac{1}{2\pi}\int_\Gamma u(z)\,[\partial^2 g(z,\zeta)/\partial n\partial\xi + \alpha(\partial g(z,\zeta)/\partial n) + \sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu(z)/\partial n)]\,ds,$$

where $\alpha, \lambda_1, \cdots, \lambda_{n-1}$ are arbitrary real constants.

Since $|\,u(z)\,| \leq 1$ in D, we have

$$(8)\quad |\,u_\xi(\zeta)\,|$$
$$\leq \tfrac{1}{2\pi}\int_\Gamma \Big|\,\partial^2 g(z,\zeta)/\partial n\partial\xi + \alpha(\partial g(z,\zeta)/\partial n) + \sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu(z)/dn)\,\Big|\,ds.$$

For any choice of the arbitrary parameters α, λ_ν, (8) yields an upper bound for $|\,u_\xi(\zeta)\,|$. In order to obtain the best possible upper bound of this type, we minimize the right-hand side of (8) with respect to the parameters

$\alpha, \lambda_1, \cdots, \lambda_{n-1}$. This is a problem in ordinary calculus, which certainly has at least one solution. Indeed, as shown by (8), the right-hand side of (8) is bounded from below; moreover, any converging subsequence of a minimizing sequence (in the variables α, λ_ν) must converge to a finite limit. If this were not the case and one or more of the variables tended to ∞ or $-\infty$, there would exist one variable, say λ_μ, such that $\limsup |\lambda_\nu/\lambda_\mu| \leq 1$ $(\nu \neq \mu)$, $\limsup |\alpha/\lambda_\mu| \leq 1$. Since we are dealing with a minimizing sequence, the right-hand side of (8) will be bounded; if we divide the integrand by $|\lambda_\mu|$, the value of the integral will tend to zero. This, however, involves a contradiction, since we may choose a subsequence for which the integrand tends to a well-defined positive function; it cannot vanish identically, since the coefficient of $\partial\omega_\mu/\partial n$ is 1.

Let now $\alpha, \lambda_1, \cdots, \lambda_{n-1}$ denote the particular values of these parameters which minimize the right-hand side of (8). With the abbreviation

$$P = \partial^2 g(z, \zeta)/\partial n \partial \xi + \alpha(\partial g(z, \zeta)/\partial n) + \sum_{\nu=1}^{n-1} \lambda_\nu (\partial \omega_\nu(z)/\partial n),$$

it follows from the minimum property that

$$(9) \qquad \int_\Gamma | P + \epsilon(\partial\omega_\nu/\partial n) | \, ds \geq \int_\Gamma | P | \, ds,$$

where ϵ is a small real parameter. Similarly,

$$(10) \qquad \int_\Gamma | P + \epsilon(\partial g(z, \zeta)/\partial n) | \, ds \geq \int_\Gamma | P | \, ds.$$

(9) is equivalent to $\epsilon \int_\Gamma | P |/P(\partial\omega_\nu/\partial n)\, ds + o(\epsilon) \geq 0$, whence, in view of the arbitrariness of the sign of ϵ,

$$(11) \qquad \int_\Gamma | P |/P(\partial\omega_\nu/\partial n)\, ds = 0.$$

In the same way, (10) leads to

$$(12) \qquad \int_\Gamma | P |/P(\partial g(z, \zeta)/\partial n)\, ds = 0.$$

The function $| P |/P$ is not defined at points at which $P = 0$. This, however, does not affect the integrals (11) and (12) since these points are finite in number (it is easy to show that the number of zeros of P on Γ cannot exceed $2n$) and we may there assign to P arbitrary values and understand the integrals in the Lebesgue sense. Another possibility is that of breaking

Γ up into arcs connecting two adjacent zeros of P and extending the definition of $|P|/P$ to the ends of these arcs by continuity. Evidently, both procedures give identical results. Whenever integrals of the type (11) and (12) will appear in the sequel, it will always be assumed, without being expressly stated each time, that the definition of the integrand has been completed by one or the other of these procedures.

We now introduce the bounded function $U(z)$ which is harmonic in D and takes the boundary values -1 or $+1$ at points at which P is positive or negative, respectively. We have $U(z) = -|P|/P$, $z \varepsilon \Gamma$, $P \neq 0$. Although the boundary values at the finite number of points at which $P = 0$ are not specified, $U(z)$ is uniquely determined by this Dirichlet problem in view of the additional assumption that $U(z)$ be bounded. $U(z)$ is easily seen to be of the form

$$U(z) = \tfrac{1}{2\pi} \int_\Gamma |P|/P(dg(z,t)/\partial n_t)\,ds_t.$$

In terms of this function $U(z)$, conditions (11) and (12) can be rewritten in the form

$$(11') \qquad \int_\Gamma U(z)\,(\partial\omega_\nu(z)/\partial n)\,ds = 0, \qquad \nu = 1, \cdots, n-1,$$

$$(12') \qquad \int_\Gamma U(z)\,(\partial g(z,\zeta)/\partial n)\,ds = 0.$$

(12′) expresses the fact that $U(\zeta) = 0$. If $V(z)$ denotes the harmonic conjugate of $U(z)$, it follows from (11′) and a manipulation similar to (5′) that $V(z)$ is free of periods about the boundary components Γ_ν, $\nu = 1, \cdots, n-1$ (and therefore also about Γ_n). $V(z)$ is only determined up to an arbitrary constant, which will be so chosen that $V(\zeta) = 0$.

The analytic function $\Phi(z) = U(z) + iV(z)$ is thus found to have the following properties: $\Phi(z)$ is regular and single-valued in D; on Γ, we have, except at a finite number of points, $|\operatorname{Re}\{\Phi(z)\}| = 1$. Since $|\operatorname{Re}\{\Phi(z)\}| \leq 1$ in D, $w = \Phi(z)$ maps D conformally onto the multiply-covered strip $-1 < \operatorname{Re}\{w\} < 1$. We further have $\Phi(\zeta) = 0$. It is easy to see that the function $\Phi(z)$ solves our extremal problem. Since the boundary values of $U(z)$ are -1 or $+1$, according as the expression within the absolute value signs on the right-hand side of (8) is positive or negative, (8) may also be written

$$u_\xi(\zeta) \leq -\tfrac{1}{2\pi} \int_\Gamma U(z)\,[\partial^2 g(z,\zeta)/\partial n\partial\xi + \alpha(\partial g(z,\zeta)/\partial n) + \sum_{\nu=1}^{n-1} \lambda_\nu(\partial\omega_\nu(z)/\partial n)]\,ds$$

whence, by (7), $u_\xi(\zeta) \leq U_\xi(\zeta)$, which proves our assertion.

We note that

$$(13)\quad U(z)P(z)$$
$$= U(z)\,[\partial^2 g(z,\zeta)/\partial n\partial\xi + \alpha(g(z,\zeta)/\partial n) + \sum_{\nu=1}^{n-1}\partial\omega_\nu(z)/\partial n] \leq 0, \quad z\,\varepsilon\,\Gamma.$$

(13) shows that $w = \Phi(z)$ yields a $(1, n)$ conformal map of D onto the strip $-1 < \mathrm{Re}\{w\} < 1$. It is well known, and easily proved, that a function of the type of $P(z)$ cannot have more than $2n$ changes of sign on Γ. (13) shows that the same is true of $U(z)$. Since each sheet of the conformal map of D given by $w = \Phi(z)$ corresponds to two changes of sign of $U(z)$ on Γ, it follows that this map cannot have more than n sheets. On the other hand, elementary topological considerations show that the number of these sheets cannot be smaller than n. It thus follows that $w = \Phi(z)$ yields a $(1, n)$ mapping of D onto the strip $-1 < \mathrm{Re}\{w\} < 1$.

The inequality (13) can be brought into a different form in which its connection with the functions of class B is more obvious. In view of (1), the function $F(z)$ of B which corresponds to the function $\Phi(z) = U(z) + iV(z)$ of B_R is given by

$$(14)\qquad F(z) = \tan(\pi/4)\Phi(z).$$

On Γ, the value of $U(z)$ is either 1 or -1, that is, $\Phi(z)$ is of the form $\Phi(z) = \pm 1 + it$, where t is a real parameter. It therefore follows from (14) by an elementary computation that $F(z)/\{1 + F^2(z)\} = \pm \frac{1}{2}\cosh t$, whence

$$(15)\qquad F(z)/\{U(z)[1 + F^2(z)]\} \geq 0, \qquad z\,\varepsilon\,\Gamma.$$

If $p(z,\zeta)$ denotes the analytic function whose real part is $g(z,\zeta)$, and $w_\nu(z)$ denotes the analytic function whose real part is $\omega_\nu(z)$, it follows from the Cauchy-Riemann equations and the fact that the real parts of both $p(z,\zeta)$ and $w_\nu(z)$ are constant on the boundary components Γ_ν, that

$$(16)\quad (\partial g(z,\zeta)/\partial n)ds = -ip'(z,\zeta)dz,\ (\partial\omega_\nu(z)/\partial n)ds = -iw'_\nu(z)dz, \quad z\,\varepsilon\,\Gamma.$$

If $P(z)$ is the function defined in (13), we have therefore

$$(17)\quad P(z)ds = -i[\partial^2 p(z,\zeta)/\partial z\partial\xi + \alpha p'(z,\zeta) + \sum_{\nu=1}^{n-1}\lambda_\nu w'_\nu(z)]dz$$
$$= R(z)dz, \quad z\,\varepsilon\,\Gamma.$$

$R(z)$ is a single-valued analytic function which is regular in $D + \Gamma$, except at ζ, where it is of the form $R(z) = -1/(z-\zeta)^2 - \alpha/(z-\zeta) +$ regular terms. Combining (13), (15), and (17), we obtain $iF(z)R(z)/\{1 + F^2(z)\}dz \geqq 0$, with the notation

$$q(z) = -R(z)/\{1+F^2(z)\} = 1/(z-\zeta)^2 + \alpha/(z-\zeta) + \text{regular terms}, \tag{18}$$

this can be written

$$-iF(z)q(z)dz \geq 0, \qquad z \varepsilon \Gamma. \tag{19}$$

The function $q(z)$ is, except for its double pole at $z=\zeta$, regular in $D+\Gamma$, since the zeros of $1+F^2(z)$ obviously coincide with the points at which $U(z)$ changes its sign; these points, in turn, coincide with the zeros of the function $P(z)$ of (13), and therefore also with those of the function $R(z)$.

Summing up, we have the following result [4]:

If $f(z) \varepsilon B$ and $\zeta \varepsilon D$, then $|f'(\zeta)| \leq F'(\zeta)$, where $F(z)$ yields a $(1, n)$ conformal map of D onto the unit circle; there exists a function $q(z)$ which is regular in $D+\Gamma$, apart from a double pole at $z=\zeta$, such that $F(z)$ and $q(z)$ are connected by the inequality (19).

It is worth pointing out that once the existence of two functions $F(z)$ and $q(z)$ which have the required properties and are connected by (19), is shown, the extremal property of $F(z)$ and a number of related results follow easily by a proper application of the residue theorem [4]. We also note that the constant α in (18) is zero; indeed, from (18), (19), the residue theorem, and the fact that $|F(z)| \equiv 1$ $(z \varepsilon \Gamma)$, it follows that

$$\alpha = \frac{1}{2\pi i}\int_\Gamma q(z)dz = \frac{1}{2\pi i}\int_\Gamma [F(z)]^* F(z)q(z)dz$$
$$= [\,\frac{1}{2\pi i}\int_\Gamma F^2(z)q(z)dz]^*,$$

and this vanishes by Cauchy's theorem, since $F(\zeta) = 0$.

3. Bounds for higher derivatives. In view of the relation (1), the problem of maximizing $|f'(\zeta)|$, $(f(z) \varepsilon B)$ and that of maximizing $|\phi'(\zeta)|$ $(\phi(z) \varepsilon B_R)$ are one and the same thing. However, this is clearly not the case if we consider the corresponding extremal problem for higher derivatives of $f(z)$ or $\phi(z)$, and we are thus faced with two different types of problems. We shall here confine ourselves to the class B_R; the class B presents some additional difficulties which seem beyond the reach of the method of this paper.

The domain of variablility of $\phi'(\zeta)$ $(\phi(z) \varepsilon B_R, \zeta \varepsilon D)$ is a circle. This, however, is not true any more in the case of higher derivatives. In order to find a characterization of all points on the boundary of the domain of variability of $\phi^{(k)}(\xi)$, we therefore have to replace the problem $|\phi^{(k)}(\zeta)| = \max.$ by the more general problem

$$\text{(20)} \qquad \mathrm{Re}\{e^{i\theta}\phi^{(k)}(\zeta)\} = \max., \qquad \phi(z) \,\varepsilon\, B_R,\ \zeta \,\varepsilon\, D,$$

where θ is an arbitrary fixed angle.

Let $q(z,\zeta) = N(z,\zeta) + iM(z,\zeta)$ denote the analytic function whose real part is the Neumann's function $N(z,\zeta)$ of D. On Γ, we have $\partial N/\partial n = -2\pi/l$, where l is the length of Γ. If $v(z)$ is harmonic in D, an application of Green's formula to the domain obtained from D by deleting from it a small circle with the center ζ yields

$$v(\zeta) = \tfrac{1}{2\pi}\int_\Gamma N(z,\zeta)\,(\partial v(z)/\partial n)\,ds - l^{-1}\int_\Gamma v(z)\,ds.$$

If $v(z)$ is the imaginary part of the analytic function $\phi(z) = u(z) + iv(z)$, this formula can be cast into a different form. Using the Cauchy-Riemann equations and integration by parts, we obtain

$$\int_\Gamma N(z,\zeta)\,(\partial v/\partial n)\,ds = -\int_\Gamma N(z,\zeta)\,(\partial u/\partial s)\,ds = -\int_\Gamma N(z,\zeta)\,du$$
$$= \int_\Gamma u\,dN(z,\zeta) = \int_\Gamma u(\partial N(z,\zeta)/\partial s)ds = -\int_\Gamma u(\partial M(z,\zeta)/\partial n)\,ds.$$

Hence,

$$\text{(21)} \qquad v(\zeta) = -\tfrac{1}{2\pi}\int_\Gamma u(z)\,(\partial M(z,\zeta)/\partial n)\,ds - l^{-1}\int_\Gamma v(z)\,ds.$$

Combining (3) and (21) we obtain

$$\mathrm{Re}\{e^{i\theta}\phi(\zeta)\} = \cos\theta u(\zeta) - \sin\theta v(\zeta) = -\tfrac{1}{2\pi}\int_\Gamma u(z)$$
$$[\cos\theta(\partial g(z,\zeta)/\partial n - \sin\theta(\partial M(z,\zeta)/\partial n)]ds + l^{-1}\sin\theta\int_\Gamma v(z)\,ds.$$

Differentiating k times with respect to $\xi(\zeta = \xi + i\eta)$ and observing (5), we arrive at the identity

$$\text{(22)} \quad \mathrm{Re}\{e^{i\theta}\phi^{(k)}(\zeta)\} = -\tfrac{1}{2\pi}\int_\Gamma u(z)\,[\cos\theta(\partial^{k+1}g(z,\zeta)/\partial\xi^k\partial n)$$
$$-\sin\theta(\partial^{k+1}M(z,\zeta)/\partial\xi^k\partial n) + \sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu/\partial n)]ds,$$

where the λ_ν are arbitrary real parameters.

The ensuing manipulation is identical with that applied to (7). Since $\phi(z)$ is in B_R, we have $|u(z)| \leq 1$, whence

$$(23)\quad \operatorname{Re}\{e^{i\theta}\phi^{(k)}(\zeta)\} \leq \frac{1}{2\pi}\int_\Gamma |\cos\theta(\partial^{k+1}g(z,\zeta)/\partial\xi^k\partial n) \\ -\sin\theta(\partial^{k+1}M(z,\zeta)/\partial\xi^k\partial n) + \sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu/\partial n)|\,ds = \frac{1}{2\pi}\int_\Gamma |P|\,ds.$$

$\int_\Gamma |P|\,ds$ is minimized with respect to the parameters λ_ν. As shown before, this leads to the conditions (11). We again define the bounded harmonic function $U(z)$ which takes the boundary values 1 at points of Γ at which P is negative, and the boundary values -1 at points at which P is positive. It follows, as before, that $U(z)$ satisfies the conditions (11′) which were shown to be equivalent to the fact that the harmonic conjugate $V(z)$ of $U(z)$ is single-valued in D. If the single-valued and regular analytic function $\Phi(z)$ is defined by $\Phi(z) = U(z) + iV(z)$, we thus have, in view of (22), (23), and the fact that $|P| = -PU$,

$$\operatorname{Re}\{e^{i\theta}\phi^{(k)}(\zeta)\} \leq \frac{1}{2\pi}\int_\Gamma |P|\,ds = -\frac{1}{2\pi}\int_\Gamma UP\,ds = \operatorname{Re}\{e^{i\theta}\Phi^{(k)}(\zeta)\},$$

that is,

$$(24)\qquad \operatorname{Re}\{e^{i\theta}\phi^{(k)}(\zeta)\} \leq \operatorname{Re}\{e^{i\theta}\Phi^{(k)}(\zeta)\}.$$

Since the function $\Phi(z)$ is in B_R, (23) shows that $\Phi(z)$ indeed solves our extremal problem.

Except for a finite number of points, the real part of $\Phi(z)$ is equal either to 1 or to -1 at all points of Γ, and it is bounded in D; it follows that $w = \Phi(z)$ maps D conformally onto the multiply-covered strip $-1 < \operatorname{Re}\{w\} < 1$. In order to determine the number of coverings, we observe that $|P| = -PU$, whence

$$(25)\quad UP\,ds \equiv U(z)[\cos\theta(\partial^{k+1}g(z,\zeta)/\partial\xi^k\partial n) \\ -\sin\theta(\partial^{k+1}M(z,\zeta)\partial\xi^k\partial n) + \sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu/\partial n)]\,ds \leq 0,$$

$z \in \Gamma$. In view of $q(z,\zeta) = N(z,\zeta) + iM(z,\zeta)$, $\partial N(z,\zeta)/\partial n = -2\pi l^{-1}$ $(z \in \Gamma)$, and the Cauchy-Riemann equations, we have, along the boundary,

$$q'(z,\zeta)\,dz = dq(z,\zeta) = dN(z,\zeta) + i\,dM(z,\zeta) \\ = (\partial N(z,\zeta)/\partial s)\,ds + i(\partial M(z,\zeta)/\partial s)\,ds \\ = -(\partial M(z,\zeta)/\partial n)\,ds + i(\partial N(z,\zeta)/\partial n)\,ds$$

$= -(\partial M(z,\zeta)/\partial n)\,ds - (2\pi i/l)\,ds$, that is,

$$(26)\qquad (\partial M(z,\zeta)/\partial n)\,ds = -q'(z,\zeta)\,dz - (2\pi i/l)\,ds, \qquad z \in \Gamma.$$

With the help of the identities (26) and (16), (25) can be brought into the form

$$(27)\qquad -IU(z)R(z)\,dz \geqq 0, \qquad z \in \Gamma,$$

where

$$R(z) = \partial^k/\partial\xi^k[\cos\theta p'(z,\zeta) + i\sin\theta q'(z,\zeta)] + \sum_{\nu=1}^{n-1}\lambda_\nu w'_\nu(z).$$

The analytic function $R(z)$ is clearly single-valued in D and its only singularity in $D+\Gamma$ is a pole at $z=\zeta$, in the neighborhood of which it has the expansion

$$(28)\qquad R(z) = -k!e^{i\theta}/(z-\zeta)^{k+1} + \text{regular terms}.$$

(27) shows that the differential $iR(z)\,dz$ is real on Γ. The total variation of $\arg\{dz\}$ on Γ is $-2\pi(n-2)$. If μ_1 denotes the number of zeros of $R(z)$ in D and μ_2 the number of its zeros on Γ, it follows therefore from (28) and the argument principle that $2\mu_1 + \mu_2 = 2(n-2+k+1) = 2(n+k-1)$. Hence, $\mu_2 \leqq 2(n+k-1)$. On the other hand, (27) shows that each change of sign of $U(z)$ on Γ necessarily coincides with a zero of $R(z)$. Since two such changes of sign correspond to one sheet of the multiply-covered strip, it follows that the maximum number of sheets is $n+k-1$.

We have thus proved

THEOREM I. *The extremal problem* (20) *is solved by a function* $w = \Phi(z)$ *which yields a* $(1, m)$ *conformal map of* D *onto the strip* $-1 < \mathrm{Re}\{w\} < 1$, *where* $m \leqq n+k-1$.

Since the domain of variability of $\phi^{(k)}(\zeta)$ $(\phi(z) \in B_R)$ is necessarily simply-connected—if $\phi_1(z)$ and $\phi_2(z)$ are in B_R, so is $\lambda\phi_1(z) + (1-\lambda)\phi_2(z)$, $0 \leqq \lambda \leqq 1$ (which shows that the domain in question is even convex)—this result may also be stated in the following form:

If $\phi(z)$ *is in* B_R *and* $\zeta \in D$, *then all boundary points of the domain of variability of* $\phi^{(k)}(\zeta)$ *are occupied by functions which yield a* $(1, m)$ *mapping of* D *onto the strip* $-1 < \mathrm{Re}\{w\} < 1$, *where* $m \leqq n+k-1$.

It is clear from the proof of Theorem I that any function solving the extremal problem (20) has necessarily the mapping properties mentioned in the statement of the theorem. This, however, does not imply that the extremal function is in all cases unique. If there are two or more different sets $\{\lambda_\nu\}$ which minimize the right-hand side of (23), this will result in two or more different extremal functions.

It is worth pointing out that the problem treated in this section also leads to a positive differential of the type (19). If $F(z)$ denotes the function of B which is connected with $\Phi(z)$ by the relation (1), then $F(z)[1+F^2(z)]^{-1}$ is real on Γ and—as shown above—its sign coincides with that of $U(z)$. If we set $R(z)/\{1+F^2(z)\} = -Q(z)$, (27) yields therefore $-iF(z)Q(z)dz \geq 0$, $z \varepsilon \Gamma$; since the zeros of $1+F^2(z)$ on Γ coincide with part of the zeros or $R(z)$, $Q(z)$ is regular in $D+\Gamma$, with the exception of a pole of order $k+1$ at $z=\zeta$.

4. Interpolation problems. The problem to be treated in this section is the following: Let $\alpha_1, \cdots, \alpha_p$ be points of D and let $a_1, \cdots, a_p$ be complex numbers such that there exists a function $\phi_0(z)$ of B_R for which $\phi_0(\alpha_i) = a_i$, $i = 1, \cdots, p$; let $B_R{}^{(\alpha)}$ denote the sub-class of B_R which consists of all functions $\phi(z)$ that satisfy the same interpolation conditions as $\phi_0(z)$ at the points α_i, and let S_k denote the domain of variability of $\phi^{(k)}(\zeta)$ ($\zeta \varepsilon D, \phi(z) \varepsilon B_R{}^{(\alpha)}$). What can be said about the functions $\phi(z)$ which occupy the boundary points of S_k? The problem thus consists in characterizing the functions which solve the extremal problem

$$\text{(29)}\quad \operatorname{Re}\{e^{i\theta}\phi^{(k)}(\zeta)\} = \text{max.}, \quad \phi(\alpha_i) = a_i \qquad (i = 1, \cdots, p), \phi(z) \varepsilon B_R.$$

If $k = 1$, it follows from (1) that this problem is identical with the corresponding interpolation problem for the class B—a generalization to multiply-connected domains of the classical Pick-Nevanlinna interpolation problem [9, 10]—which was treated, by different methods, in [4].

We write $a_i = b_i + ic_i$ (b_i, c_i real) and normalize the functions $\phi(z)$ by the requirement

$$\text{(30)}\qquad \sum_{i=1}^{p} c_i = 0.$$

This is possible since the addition of an imaginary constant to $\phi(z)$ does not alter the fact that $\phi(z)$ is in B_R. Utilizing (21), (22), and the fact that $\phi(\alpha_i) = u(\alpha_i) + iv(\alpha_i) = b_i + ic_i$, we obtain the representation

$$\begin{aligned}\text{(31)}\quad \operatorname{Re}\{e^{i\theta}\phi^{(k)}(\zeta)\} = -\frac{1}{2\pi}\int_{\Gamma} u(z)\,[\cos\theta\,(\partial^{k+1}g(z,\zeta)/\partial\xi^k\partial n) \\ -\sin\theta\,(\partial^{k+1}M(z,\zeta)/\partial\xi^k\partial n) + \sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu/\partial n) + \sum_{i=1}^{p}(\sigma_i(\partial g(z,\alpha_i)/\partial n) \\ + \tau_i(\partial M(z,\alpha_i)/\partial n))]ds - \sum_{i=1}^{p}(\sigma_i b_i + \tau_i c_i),\end{aligned}$$

where the λ_ν, σ_i are arbitrary real constants, and the otherwise arbitrary real constants τ_i satisfy the condition

$$\sum_{i=1}^{p} \tau_i = 0 \tag{30'}$$

(30′) has the effect of cancelling the contributions due to the second term on the right-hand side of (21).

Since $|u(z)| \leq 1$, it follows from (31) that

$$\begin{aligned}(32)\quad & \operatorname{Re}\{e^{i\theta}\phi^{(k)}(z)\} \\ & \leq \tfrac{1}{2\pi}\int_\Gamma \Big| \cos\theta(\partial^{k+1}g(z,\zeta)/\partial\xi^k\partial n) - \sin\theta(\partial^{k+1}M(z,\zeta)/\partial\xi^k\partial n) \\ & \quad + \sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu/\partial n) + \sum_{i=1}^{p}(\sigma_i(\partial g(z,\alpha_i)/\partial n) + \tau_i(\partial M(z,\alpha_i)/\partial n))\Big|\, ds \\ & \quad - \sum_{i=1}^{p}(\sigma_i b_i + \tau_i c_i) = \tfrac{1}{2\pi}\int_\Gamma |P|\, ds - \sum_{i=1}^{p}(\sigma_i b_i + \tau_i c_i).\end{aligned}$$

As in the previous sections, we now proceed to minimize the right-hand side of (32) by a suitable choice of the constants λ_ν, σ_i, τ_i. The existence of the minimum is assured; as shown by (31) and (32), the value of $\operatorname{Re}\{e^{i\theta}\phi_0^{(k)}(\zeta)\}$—where $\phi_0(z)$ is any function of $B_R^{(\alpha)}$— furnishes a lower bound for the quantity in question, and the existence of a minimizing set λ_ν, σ_i, τ_i follows in the same way as in a similar case in Section 2. Suppose now that the values λ_ν, σ_i, τ_i, have been so chosen as to yield the minimum of the right-hand side of (32) under the condition (30′). As in Section 2, a small variation of the λ_ν yields the conditions

$$\int_\Gamma |P|/P(\partial\omega_\nu/\partial n)\,ds = 0, \qquad \nu = 1,\cdots,n-1. \tag{33}$$

Next, we vary the σ_i. Replacing σ_j by $\sigma_j + \epsilon$, we obtain

$$\begin{aligned}\tfrac{1}{2\pi}\int_\Gamma |P|\,ds + \epsilon/2\pi\int_\Gamma |P|/P(\partial g(z,\alpha_j)/\partial n)ds + o(\epsilon) - \sum_{i=1}^{p}(\sigma_i b_i + \tau_i c_i) - \epsilon b_j \\ \geq \tfrac{1}{2\pi}\int_\Gamma |P|\,ds - \sum_{i=1}^{p}(\sigma_i b_i + \tau_i c_i),\end{aligned}$$

whence, in view of the arbitrariness of the sign of ϵ,

$$\tfrac{1}{2\pi}\int_\Gamma |P|/P(\partial g(z,\alpha_j)/\partial n)\,ds = b_j. \tag{34}$$

In varying the τ_i, we have to take into account the condition (30′). Replacing, accordingly, τ_j by $\tau_j + \epsilon$ and τ_q $(q \neq j)$ by $\tau_q - \epsilon$, and observing the arbitrariness of the sign of ϵ, we obtain in the same manner as before

$$(35) \quad \frac{1}{2\pi}\int_\Gamma |P|/P(\partial M(z,\alpha_j)/\partial n)\,ds - \frac{1}{2\pi}\int_\Gamma |P|/P(\partial M(z,\alpha_q)/\partial n)\,ds = c_j - c_q.$$

We now introduce the bounded harmonic function $U(z)$ whose boundary values on Γ are -1 or $+1$, according as P is positive or negative, respectively, that is,

$$(36) \qquad U(z) = -|P|/P, \quad z \,\varepsilon\, \Gamma(P \neq 0).$$

In terms of this function, the relations (33), (34), (35) take the form

$$(33') \qquad \int_\Gamma U(z)\left(\partial\omega_\nu(z)\partial n\right)ds = 0$$

$$(34') \qquad -\frac{1}{2\pi}\int_\Gamma U(z)\left(\partial g(z,\alpha_j)/\partial n\right)ds = b_j$$

$$(35') \quad -\frac{1}{2\pi}\int_\Gamma U(z)\left(\partial M(z,\alpha_j)/\partial n\right)ds + \frac{1}{2\pi}\int_\Gamma U(z)\left(\partial M(z,\alpha_q)/\partial n\right)ds = c_j - c_q.$$

As shown by (5′), it follows from (33′) that the harmonic conjugate $V(z)$ of $U(z)$ is single-valued in D. The same is therefore true of the analytic function $\phi(z) = U(z) + iV(z)$, which thus belongs to B_R. (34′) and (35′) are—in view of (3) and (21)—equivalent to

$$(37) \qquad U(\alpha_j) = b_j$$

and $V(\alpha_j) - V(\alpha_q) = c_j - c_q$. In view of (30), it follows from the last formula that

$$(38) \qquad pV(\alpha_j) - \sum_{q=1}^{p} V(\alpha_q) = pc_j\,.$$

The harmonic conjugate $V(z)$ of $U(z)$ is only determined up to an arbitrary constant. If this constant is so chosen that $\sum_{q=1}^{p} V(\alpha_q) = 0$, (38) becomes $V(\alpha_j) = c_j$. Combining this with (37), we finally obtain

$$\Phi(\alpha_i) = U(\alpha_j) + iV(\alpha_j) = b_j + ic_j = a_j,$$

which shows that the function $\Phi(z)$ satisfies the interpolation conditions of our problem; in view of its other properties found further above, it is therefore in $B_R^{(\alpha)}$.

The function $\Phi(z)$ solves our extremal problem; indeed, as shown by (31), (32), and (36), we have

$$\operatorname{Re}\{e^{i\theta}\phi^{(k)}(\zeta)\} \leq \tfrac{1}{2\pi}\int_\Gamma |P|\,ds - \sum_{i=1}^{p}(\sigma_i b_i + \tau_i c_i)$$

$$= -\tfrac{1}{2\pi}\int_\Gamma U(z)P\,ds - \sum_{i=1}^{p}(\sigma_i b_i + \tau_i c_i) = \operatorname{Re}\{e^{i\theta}\Phi^{(k)}(\zeta)\}.$$

Before we state our results in the form of a theorem, we show that our problem is again closely connected with a certain positive differential. By (16), (26), (32), and (30′),

$$(39)\quad P\,ds = -i\{\partial^k/\partial\xi^k[\cos\theta p'(z,\zeta) + i\sin\theta q'(z,\zeta)] + \sum_{\nu=1}^{n-1}\lambda_\nu w'_\nu(z)$$
$$+ \sum_{i=1}^{p}[\sigma_i p'(z,\alpha_i) - i\tau_i q'(z,\alpha_i)]\}dz \equiv iR(z)\,dz, \quad z \in \Gamma,$$

where the single-valued function $R(z)$ is regular in $D + \Gamma$, with the exception of the point $z = \zeta$ where it has a pole of order $k + 1$, and of the points $z = \alpha_i$, $i = 1, \cdots, p$, where it has single poles. Comparison of (39) and (36) shows that we have $-iU(z)R(z)dz \geq 0$, $z \in \Gamma$. If we introduce the bounded function $F(z)$ which is associated with $\phi(z)$ by means of (1), it follows from the relation (15), which obviously also holds for the functions $U(z)$ and $F(z)$ of this section, that

$$-iF(z)R(z)/\{1 + F^2(z)\}dz \geq 0, \qquad z \in \Gamma,$$

or, with the notation $Q(z) = R(z)/\{1 + F^2(z)\}$,

$$(40)\qquad -iF(z)Q(z)\,dz \geq 0, \qquad z \in \Gamma.$$

The function $Q(z)$ has the same types of singularities as $R(z)$; the zeros of $1 + F^2(z)$ on Γ do not give rise to poles of $Q(z)$, since—by virtue of (1), (36) and (39)—they clearly coincide with part of the zeros of $R(z)$. (40) permits an easy counting of the number of sheets of the conformal image of D by means of $F(z)$—or, what amounts to the same thing, by means of $\Phi(z)$. The total number of poles of $Q(z)$ is $k + p + 1$, where p is the number of interpolation points; the total variation of $\arg\{dz\}$ along Γ is $-2\pi(n - 2)$. Since the argument of the left-hand side of (40) is constant on Γ, it therefore follows from the argument principle that the total number of zeros of

$F(z)Q(z)$ in $D+\Gamma$ (zeros on Γ to be counted with half their multiplicity) is $k+p+n-1$. Hence, the total number of zeros of $F(z)$ in D is smaller than or equal to $k+p+n-1$. Since $|F(z)|\equiv 1$ on Γ, this number is identical with the number of times the unit circle is covered by the values of $F(z)$ for $z\,\varepsilon\, D$.

We have thus proved the following theorem:

THEOREM II. *Let* $\alpha_1,\cdots,\alpha_p$ *be points of* D *and let* $a_1,\cdots,a_p$ *be complex numbers—which, for convenience, are normalized by the requirement* $\mathrm{Im}\{a_1+\cdots+a_p\}=0$*—such that there exists a function* $\phi_0(z)$ *of* B_R *for which* $\phi_0(\alpha_i)=a_i$, $i=1,\cdots,p$; *if* $B_R^{(a)}$ *denotes the sub-class of* B_R *which consists of all functions* $\phi(z)$ *which satisfy the same interpolation conditions as* $\phi_0(z)$ *at the points* α_i, *then*

$$\mathrm{Re}\{e^{i\theta}\phi^{(k)}(\zeta)\}\leqq\mathrm{Re}\{e^{i\theta}\Phi^{(k)}(\zeta)\},\quad \zeta\,\varepsilon\, D,\quad 0\leqq\theta<2\pi,\quad \Phi(z)\subset B_R^{(a)},$$

where $w=\Phi(z)$ *yields a* $(1,m)$ *conformal mapping of* D *onto the strip* $-1<\mathrm{Re}\{w\}<1$, *and* $m\leq k+p+n-1$. *The bounded function* $F(z)$, *which is associated with* $\Phi(z)$ *by means of* (1), *satisfies the relation* (40), *where the function* $Q(z)$ *is single-valued and regular in* D, *with the exception of a pole of order* $k+1$ *at* $z=\zeta$ *and of* p *simple poles at the points* α_i $(i=1,\cdots,p)$.

We remark that by exactly the same method of proof we may treat the more general problem

$$\mathrm{Re}\{\sum_{j=1}^{v}\rho_j\phi^{(k_j)}(\zeta_j)\}=\max.,\quad \phi(z)\,\varepsilon\, B_R^{(a)},\quad \zeta_j\,\varepsilon\, D,\quad j=1,\cdots,r,$$

where the ρ_j are given complex numbers. The function solving this problem will again have the same general characteristics as $\Phi(z)$; the only difference being that the number of sheets of the extremal mapping may now increase to $p+n+k_1+\cdots+k_r-1$.

It has already been pointed out above that, in the case $k=1$, the interpolation problems for the classes B and B_R, respectively, are identical. In this connection it is worth noting that the extremal property of the function $F(z)$ is an immediate consequence of the inequality (40). We recall that the function $Q(z)$ in (40) has simple poles—with the residues η_i, say—at the points α_i and a double pole at $z=\zeta$; if for the sake of simplicity, we add the interpolation condition $\phi(\zeta)=0$—or, what amounts to the same thing, $f(\zeta)=0$—the principal part of $Q(z)$ at $z=\zeta$ is easily found to be of the form $e^{i\theta}(z-\zeta)^{-2}$. Let now $f(z)$ be in B and let $f(\alpha_i)=A_i$, $i=1,\cdots,p$, $f(\zeta)=0$. The function $F(z)$ in (40) satisfies the same interpolation conditions, and we have $|F(z)|\equiv 1$, $z\,\varepsilon\,\Gamma$. Using this fact, the residue theorem, and (40), we obtain

$$\begin{aligned}\operatorname{Re}\{e^{i\theta}f'(\zeta)+\sum_{i=1}^{p}\eta_iA_i\}&=\operatorname{Re}\{\tfrac{1}{2\pi i}\int_\Gamma Q(z)f(z)\,dz\}\\ &\leq\tfrac{1}{2\pi}\int_\Gamma|Q(z)\,dz|=\tfrac{1}{2\pi}\int_\Gamma|F(z)Q(z)\,dz|=\tfrac{1}{2\pi i}\int_\Gamma F(z)Q(z)\,dz\\ &=\operatorname{Re}\{\tfrac{1}{2\pi i}\int_\Gamma F(z)Q(z)\,dz\}=\operatorname{Re}\{e^{i\theta}F'(\zeta)+\sum_{i=1}^{p}\eta_iA_i\},\end{aligned}$$

whence $\operatorname{Re}\{e^{i\theta}f'(\zeta)\}\leq\operatorname{Re}\{e^{i\theta}F'(\zeta)\}$.

5. The radius of univalence. It was shown by Landau [7] that a function $f(z)$ which is regular and bounded in the unit circle and satisfies $|f'(0)|=a$, $0<a<1$, yields a schlicht mapping of the circle $|z|<\rho=a[1+(1-a^2)^{\frac{1}{2}}]^{-1}$. The function $f(z)$ with the smallest radius of univalence—for which this estimate is sharp—yields a $(1,2)$ mapping of the unit circle upon itself. The objective of this section is to generalize this result to functions of class B which are defined in an arbitrary multiply-connected domain. Again, we shall carry through the proofs for functions of the class B_R; obviously, the problems for the classes B and B_R, respectively, are identical.

A further remark is appropriate. In the case in which D reduces to the unit circle and the normalization condition refers to its center, it is natural to consider regions of univalence which are concentric circles. In the case of a general multiply-connected domain, however, it might be desirable to consider regions of univalence of a different type. For example, the concentric circles of the unit circle problem may be generalized in a natural way to regions bounded by level curves of the Green's function of D with its singularity at the point of normalization. The results of this section will include these more general cases.

Let now $\phi(z)$ be a function of B_R, and let $\phi'(\zeta)=A=a+ib$ $(\zeta\,\varepsilon\,D)$ be given; A, of course, has to be an admissible value for $\phi'(\zeta)$ in accordance with the result proved in Section 2. If z_1 and z_2 are two different points in D and $0\leq\theta<2\pi$, we have, by (3), (5), and (21),

$$(41)\quad \operatorname{Re}\{e^{i\theta}[\phi(z_1)-\phi(z_2)]\}$$

$$=-\tfrac{1}{2\pi}\int_\Gamma u(z)\{\cos\theta[\partial g(z,z_1)/\partial n-\partial g(z,z_2)/\partial n]$$

$$-\sin\theta[\partial M(z,z_1)/\partial n-\partial M(z,z_2)/\partial n]+\sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu/\partial n)\}ds,$$

where $\phi(z)=u(z)+iv(z)$, and the λ_ν are arbitrary real parameters. In order to carry the condition that $\phi'(\zeta)=A$, we observe that, in view of (3),

$$A=a+ib=\phi'(\zeta)=\partial u(\zeta)/\partial\xi-i\partial u(\zeta)/\partial\eta$$

$$=-\tfrac{1}{2\pi}\int_\Gamma u(z)[\partial^2 g(z,\zeta)/\partial\xi\partial n-i(\partial^2 g(z,\zeta)/\partial\eta\partial n)ds.$$

(41) can therefore be replaced by

$$(42)\quad \operatorname{Re}\{e^{i\theta}[\phi(z_1)-\phi(z_2)]\}=-\tfrac{1}{2\pi}\int_\Gamma u(z)[\cos\theta(\partial g(z;z_1,z_2)/\partial n)$$

$$-\sin\theta(\partial M(z;z_1,z_2)/\partial n)+\sum_{\nu=1}^{n-1}\lambda_\nu\ \partial\omega_\nu/\partial n)+\alpha(\partial^2 g(z,\zeta)/\partial\xi\partial n)$$

$$+\beta(\partial^2 g(z,\zeta)/\partial\xi\partial n)ds]-\alpha a+\beta b,$$

where α and β are arbitrary real parameters, and the meaning of the abbreviations $g(z;z_1,z_2)$ and $M(z;z_1,z_2)$ is obvious. Since $|u(z)|\leq 1$, it follows from (42) that

$$(43)\quad -|\phi(z_1)-\phi(z_2)|<\operatorname{Re}\{e^{i\theta}[\phi(z_1)-\phi(z_2)]\}$$

$$\leq\tfrac{1}{2\pi}\int_\Gamma|\cos\theta(\partial g(z;z_1,z_2)/\partial n)-\sin\theta(\partial M(z;z_1,z_2)/\partial n)$$

$$+\sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu/\partial n)+\alpha(\partial^2 g(z,\zeta)/\partial\xi\partial n)+\beta(\partial^2 g(z,\zeta)/\partial\eta\partial n)|ds$$

$$-\alpha a+\beta b$$

or, with an obvious abbreviation,

$$(43')\qquad -|\phi(z_1)-\phi(z_2)|\leq\tfrac{1}{2\pi}\int_\Gamma|P|\,ds-\alpha a+\beta b.$$

As in similar cases further above, we now minimize the right-hand side of (43) by a suitable choice of the arbitrary parameters $\lambda_1,\cdots,\lambda_{n-1},\alpha,\beta,\theta$;

the existence of the minimum follows in the same elementary way as before. We assume that in (43)—or (43′)—the parameters have already been so chosen that they yield the minimum of the expression concerned, and we define the bounded harmonic function $U(z)$ by the boundary values

$$U(z) = -|P|/P, \quad z \in \Gamma, \quad P \neq 0, \tag{44}$$

that is, $U(z)$ takes the boundary values -1 or $+1$, according as P is positive or negative, respectively. By an elementary computation, which is identical with those carried out in earlier similar cases and which we therefore omit, the necessary conditions for the existence of the minimum are found to be

$$\int_\Gamma U(z)\,(\partial\omega_\nu(z)/\partial n)\,ds = 0, \tag{45}$$

$$-\tfrac{1}{2\pi}\int_\Gamma U(z)\,(\partial^2 g(z,\zeta)\,\partial\xi\partial n)\,ds = a, \tag{46}$$

$$-\tfrac{1}{2\pi}\int_\Gamma U(z)\,(\partial^2 g(z,\zeta)\,\partial\eta\partial n)\,ds = -b,$$

$$-\tfrac{1}{2\pi}\int_\Gamma U(z)\,[\sin\theta(\partial g(z;z_1,z_2)/\partial n) + \cos\theta(\partial M(z;z_1,z_2)/\partial n)]\,ds = 0. \tag{47}$$

As shown in Section 2, (45) is equivalent to the fact that the harmonic conjugate $V(z)$ of $U(z)$—and therefore also the analytic function $\Phi(z) = U(z) + iV(z)$—is single-valued in D. (46) shows, in view of (3), that $\partial U/\partial x$ and $\partial U/\partial y$ have the values a and $-b$, respectively, at the point $z = \zeta$. By the Cauchy-Riemann equations, it follows that $\Phi'(\zeta) = a + bi = A$. The condition (47) is found to be equivalent to

$$\operatorname{Im}\{e^{i\theta}[\Phi(z_1) - \Phi(z_2)]\} = 0, \tag{48}$$

if we recall the definitions of $g(z;z_1,z_2)$ and $M(z;z_1,z_2)$, and use (3) and (21). We have thus shown that for the "extremal angle" θ, the expression $e^{i\theta}[\phi(z_1) - \phi(z_2)]$ is real. For our purposes, however, it is essential to show that this expression is moreover non-positive. This is done by considering the second variation of the right-hand side of (43) if θ is replaced by $\theta + \epsilon$. If the varied value of P is denoted by P^*, an elementary computation shows that

$$\int_\Gamma |P^*|\,ds = \int_\Gamma |P|\,ds - \epsilon \int_\Gamma |P|/P[\sin\theta(\partial g(z;z_1,z_2)/\partial n) + \cos\theta(\partial M(z;z_1,z_2)/\partial n)]\,ds$$
$$-\tfrac{1}{2}\epsilon^2 \int_\Gamma |P|/P[\cos\theta(\partial g(z;z_1,z_2)/\partial n) - \sin\theta(\partial M(z;z_1,z_2)/\partial n)]\,ds + o(\epsilon^2).$$

The first variation vanishes by (47). Since $\int_\Gamma |P^*|\,ds \geqq \int_\Gamma |P|\,ds$, the second variation is non-negative. In view of (44), this is equivalent to

$$\int_\Gamma U(z)\,[\cos\theta(\partial g(z;z_1,z_2)/\partial n) - \sin\theta(\partial M(z;z_1,z_2)/\partial n)\,ds \geqq 0,$$

whence, by (3) and (21), $\operatorname{Re}\{e^{i\theta}[\Phi(z_1) - \Phi(z_2)]\} \leqq 0$. Combining this with (48), we obtain

$$e^{i\theta}[\Phi(z_1) - \Phi(z_2)] = -|\Phi(z_1) - \Phi(z_2)|. \tag{49}$$

On the other hand, it follows from (43′), (44) and (41) that

$$-|\phi(z_1) - \phi(z_2)| \leqq \operatorname{Re}\{e^{i\theta}[\Phi(z_1) - \Phi(z_2)]\}.$$

Combining this with (49), we finally obtain

$$-|\phi(z_1) - \phi(z_2)| \leqq -|\Phi(z_1) - \Phi(z_2)|.$$

We thus have the following result:

THEOREM III. *Let $\phi(z)\ \varepsilon\ B_R$, $\phi'(\zeta) = A$, $\zeta\ \varepsilon\ D$, $z_1\ \varepsilon\ D$, $z_2\ \varepsilon\ D$; then*

$$|\phi(z_1) - \phi(z_2)| \geqq |\Phi(z_2) - \Phi(z_2)|, \tag{50}$$

where $\Phi(z)$ is likewise in B_R and $\Phi'(\zeta) = A$; $w = \Phi(z)$ yields a $(1, m)$ conformal map of D onto the strip $-1 < \operatorname{Re}\{w\} < 1$, where $n \leqq m \leqq n + 2$.

The limitations for m follow, in the same manner as on previous occasions, from (44) and the easily demonstrable fact that the function P cannot have more than $2(n + 2)$ changes of sign on Γ.

An immediate application of Theorem III is

THEOREM IV. *Let $f(z)\ \varepsilon\ B$ and $f'(\zeta) = A\ (\zeta\ \varepsilon\ D)$; let ρ be the largest number such that all functions $\phi(z)$ with these properties are univalent in $|z - \zeta| < \rho$. Then there exists a function $F(z)\ \varepsilon\ B$, with $F'(\zeta) = A$ which* *yields a* $(1, m)$ *map* $(n \leq m \leq n + 2)$ *of* D *onto the unit circle, such that* $\Phi(z)$ *is not univalent in any circle* $|z - \zeta| < \rho'$ *with* $\rho' > \rho$.

Remark. If D is the unit circle and ζ its center, the function $F(z)$ of Theorem IV yields a $(1, m)$ mapping of the unit circle onto itself, where $m \leq 3$. According to the theorem of Landau quoted above, $m = 2$. Whether it is always true that $m \leq n + 1$, or whether this is a special property of the circle, is an open question.

Proof. In view of the schlicht transformation (1), Theorem IV is equivalent to the corresponding result for functions $\phi(z)$ of B_R with a given derivative A at $z = \zeta$. It follows from Landau's theorem, or else by a standard argument using the fact that this sub-class $B_R{}^A$ of B_R is compact, that there exist values $\rho > 0$ such that all such functions $\phi(z)$ are univalent for $|z - \zeta| < \rho$. Let now ρ be the largest value with this property, i. e. let ρ be the radius of univalence of $B_R{}^A$. A compactness argument shows that there exists a function of $B_R{}^A$, say $\phi_0(z)$, which is univalent in $|z - \zeta| < \rho$ but in no larger circle about ζ. Clearly, there are two possibilities: either there are two different points z_1 and z_2 on the circumference $|z - \zeta| = \rho$ for which $\phi_0(z_1) = \phi_0(z_2)$, or else there exists a point z_3 on this circumference at which $\phi'_0(z_0) = 0$. In the first case, Theorem III shows that there exists a function $\Phi(z)$ in $B_R{}^A$ of the type described in the statement of Theorem IV, such that also $\Phi(z_1) = \Phi(z_2)$. It follows that $\Phi(z)$ cannot be univalent in a circle about ζ of radius larger than ρ. In the second case, we replace (50) by

$$|\phi(z_1) - \phi(z_2)|/|z_1 - z_2| \geq |\Phi(z_1) - \Phi(z_2)|/|z_1 - z_2|$$

and let both z_1 and z_2 tend to z_3. This leads to

$$|\phi'(z_3)| \geq |\Phi'_0(z_3)|, \tag{51}$$

where $\Phi_0(z)$ is the limit of a sequence of functions, all of which map D onto strips whose maximum number of sheets is $n + 2$. Since the class of these mappings is compact, $\Phi_0(z)$ yields a mapping of the same type. (51) shows that if there is a function $\phi_0(z)$ of $B_R{}^A$ for which $\phi'_0(z_3) = 0$,the same is true of a function $\Phi_0(z)$ with the properties described in the statement of Theorem IV. This completes the proof.

The fact that the regions of univalence are circles was hardly used in the proof of Theorem IV. All we used was the fact that these circles form a continuous family of domains D_ρ depending on a positive parameter ρ, which grows monotonically with ρ and, for $\rho = 0$, shrinks to the point $z = \zeta$. Theorem IV is therefore also true for other regions of univalence which form such a family. For example, we might choose the domains D_ρ to be the regions in which $g(z, \zeta) > \rho^{-1}$, that is, regions which are bounded by the level curves of the Green's function. As already said, this would be the natural generalization of the circle of univalence defined in the case of bounded functions in the unit circle.

We finally remark that, in addition to the requirement $\phi'(\zeta) = A$, the class B_R may be restricted by further conditions of an interpolatory character, and we may then ask for the exact radius of univalence of this restricted class. Combining the procedure leading to Theorem IV with some of the special features of the proof of Theorem III, one can derive the corresponding result without difficulty. If the number of interpolation conditions is k—a given value of the first derivative counts as two conditions, etc.—then the function $w = \Phi(z)$ which, within this restricted class, has the smallest radius of univalence will yield a $(1, m)$ mapping of D onto the strip $-1 < \operatorname{Re}\{w\} < 1$, where $m \leq n + k + 2$.

6. The radius of starlikeness and the radius of convexity. The radius of starlikeness $\rho_s(\zeta)$ of an analytic function $f(z)$ with respect to the point ζ, at which $f(\zeta) = 0$, is defined as the radius of the largest circle about ζ which is mapped by $f(z)$ onto a schlicht domain that is starlike with respect to the origin. Similarly, the radius of convexity $\rho_c(\zeta)$ is defined as the radius of the largest circle about ζ which is mapped by $f(z)$ onto a convex schlicht domain. In the case of the family of bounded functions in the unit circle with a given derivative at the origin, both the function with the smallest $\rho_s(0)$ and that with the smallest $\rho_c(0)$ are identical with the function with the smallest radius of univalence, mentioned at the beginning of the preceding section. The result for $\rho_s(0)$ is due to Dieudonné [3] and that for $\rho_c(0)$ to Cacridis-Theodorakopulos [2].

The method of the preceding section requires but little modification to yield the corresponding results regarding the radius of starlikeness and the radius of convexity for functions of B_R. In the case of $\rho_s(\zeta)$, we set up an expression for $\operatorname{Re}\{e^{i\theta}[\phi(z_1) - \alpha\phi(z_2)]\}$—where $\alpha > 0$ is an arbitrary parameter—corresponding to (42). The ensuing procedure is then the same as that applied to (42); it finally leads to the inequality

$$|\phi(z_1) - \alpha\phi(z_2)| \geqq |\Phi(z_1) - \alpha\Phi(z_2)| \tag{52}$$

—corresponding to (50)—where $w = \Phi(z)$ again maps D onto the m times covered strip $-1 < \operatorname{Re}\{w\} < 1$, and $m \leq n + 2$. Let now $B_R{}^A$ denote the class of functions $\phi(z)$ of B_R for which $\phi(\zeta) = 0$ and $\phi'(\zeta) = A$. If $\rho_s(\zeta)$ denotes the radius of starlikeness of this class with respect to $z = \zeta$, then there exist a function $\Phi_0(z)$ of $B_R{}^A$, and two points z_1 and z_2 on the circumference $| z - \zeta | = \rho_s(\zeta) + \epsilon (\epsilon > 0)$, such that $\phi_0(z_1) - \alpha\phi_0(z_2) = 0$ for some positive α. By (52), there also exists a function $\Phi(z)$ with the above mentioned properties for which $\Phi(z_1) - \alpha\Phi(z_2) = 0$ and which, therefore, likewise maps $| z - \zeta | < \rho_c(\zeta) + \epsilon$ onto a non-starlike domain. Letting $\epsilon \rightarrow 0$ and observing that the class of these functions $\Phi(z)$ is compact, we obtain

THEOREM V. *Let $B_R{}^A$ be the class of functions $\phi(z)$ of B_R for which $\phi(\zeta) = 0$ and $\phi'(\zeta) = A$, and let $\rho_s(\zeta)$ be the radius of starlikeness of $B_R{}^A$ with respect to ζ. Then $\rho_s(\zeta)$ is the exact radius of starlikeness of a function $w = \Phi(z)$ of $B_R{}^A$ which yields a $(1, m)$ mapping $(m \leq n + 2)$ of D onto the strip $-1 < \operatorname{Re}\{w\} < 1$.*

The result concerning the radius of convexity follows in exactly the same way, the only difference being that the expression $\phi(z_1) - \alpha\phi(z_2)$ is now replaced by

$$\lambda\phi(z_1) + (1 - \lambda)\phi(z_2) - \phi(z_3), \qquad 0 < \lambda < 1.$$

The vanishing of this expression for a triple of distinct points z_1, z_2, z_3 on a circle about $z = \zeta$, and for a certain λ, is identical with the failure of the convexity property. As a result of the fact that we have to carry the additional point z_3 in (41) and the ensuing manipulation, the maximum number of sheets of the extremal map is raised to $n + 3$. Since the proof presents no new features, we confine ourselves to the statement of the result.

THEOREM VI. *Let $B_R{}^A$ be the class of functions defined in the statement of Theorem V, and let $\rho_c(\zeta)$ be the radius of convexity of $B_R{}^A$ with respect to $z = \zeta$. Then $\rho_c(\zeta)$ is the exact radius of convexity of a function $w = \Phi(z)$ of $B_R{}^A$ which yields a $(1, m)$ mapping $(m \leq n + 3)$ of D onto the strip $-1 < \operatorname{Re}\{w\} < 1$.*

It hardly needs to be pointed out that Theorems V and VI do not solve the corresponding problems for the class B; obviously, the transformation (1) does not preserve the properties of starlikeness and convexity.

7. Other problems. By suitable modifications of the main method of proof of the preceding sections, a great variety of related problems can be solved. In this section, we shall treat two of these problems by way of illustration of the general principles involved.

The first of these problems concerns functions $\phi(z)$ which are regular and single-valued in D and for which

$$-b_\mu \leq \limsup_{z \to \Gamma_\mu} \operatorname{Re}\{\phi(z)\} - a_\mu \leq b_\mu,$$

where $a_1, \cdots, a_n$ are real constants, and $b_1, \cdots, b_n$ are positive constants. This class is a generaliaztion of the class B_R discussed before, and all results concerning functions of B_R which were obtained in the foregoing can be extended to this wider class. As an example, we prove

THEOREM VII. *Let $\phi(z)$ be regular and single-valued in D, and let*

$$(53) \qquad -b_\mu \leq \limsup_{z \to \Gamma_\mu} \operatorname{Re}\{\phi(z)\} - a_\mu \leq b_\mu, \qquad b_\mu > 0,$$

where $\Gamma_1, \cdots, \Gamma_n$ are the boundary components of D and, apart from the conditions $b_\mu > 0$, the a_μ and b_μ are arbitrary real constants. Then

$$(54) \qquad \operatorname{Re}\{\phi'(\zeta)\} \leq \operatorname{Re}\{\Phi'(\zeta)\}, \qquad \zeta \varepsilon D,$$

where $\Phi(z)$ is regular and single-valued in D and satisfies

$$(55) \qquad |\operatorname{Re}\{\Phi(z)\} - a_\mu| = b_\mu, \qquad z \varepsilon \Gamma_\mu,$$

with the possible exception of not more than $2n$ points at which the boundary values of $\operatorname{Re}\{\Phi(z)\}$ may not be defined. No value is taken by $\Phi(z)$ in D more than n times.

Proof. By Fatou's theorem, the limits of $u(z) = \operatorname{Re}\{\Phi(z)\}$ exist almost everywhere on Γ. Green's formula is valid with these boundary values of $u(z)$, since we may approximate Γ by a sequence of level curves of the Green's function and apply the Lebesgue theorem to the resulting sequence of integrals. If $u(z)$ $(z \varepsilon \Gamma)$ denotes these boundary values, it follows—as shown before—from Green's formula and the fact that $\phi(z)$ is single-valued in D that

$$\operatorname{Re}\{\phi'(\zeta)\} = -\tfrac{1}{2\pi} \int_\Gamma u(z)\,[\partial^2 g(z,\zeta)/\partial\xi\partial n + \sum_{\nu=1}^{n-1} \lambda_\nu (\partial\omega_\nu/\partial n)]\,ds,$$

where the λ_ν are arbitrary real parameters. This formula can also be written in the form

$$(56)\quad \operatorname{Re}\{\phi'(\zeta)\} = -\sum_{\mu=1}^{n} \tfrac{1}{2\pi} \int_{\Gamma_\mu} [u(z) - a_\mu]\,[\partial^2 g(z,\zeta)/\partial\xi\partial n + \sum_{\nu=1}^{n-1} \lambda_\nu(\partial\omega_\nu/\partial n)]\,ds$$
$$- \sum_{\mu=1}^{n-1} (a_\mu/2\pi) \int_{\Gamma_\mu} [\partial^2 g(z,\zeta)/\partial\xi\partial n + \sum_{\nu=1}^{n-1} \lambda_\nu(\partial\omega_\nu/\partial n)]\,ds.$$

It follows from (53) that

$$(57)\quad \operatorname{Re}\{\phi'(\zeta)\} \leq \sum_{\mu=1}^{n} (b_\mu/2\pi) \int_{\Gamma_\mu} |\,\partial^2 g(z,\zeta)/\partial\xi\partial n + \sum_{\nu=1}^{n-1} \lambda_\nu(\partial\omega_\nu/\partial n)|\,ds$$
$$- \sum_{\mu=1}^{n} (a_\mu/2\pi) \int_{\Gamma_\mu} [\partial^2 g(z,\zeta)/\partial\xi\partial n + \sum_{\nu=1}^{n-1} \lambda_\nu(\partial\omega_\nu/\partial n)]\,ds.$$

We now minimize the right-hand side of (57) with respect to the parameters λ_ν. Differentiating the right-hand side of (57) with respect to λ_ν, the necessary conditions to be satisfied in the case of the minimum are found to be

$$(58)\quad \sum_{\mu=1}^{n} (b_\mu/2\pi) \int_{\Gamma_\mu} |\,P\,|/P\,(\partial\omega_\nu/\partial n)\,ds - \sum_{\mu=1}^{n} (a_\mu/2\pi) \int_{\Gamma_\mu} (\partial\omega_\nu/\partial n)\,ds = 0,$$
$$\nu = 1, \cdots, n-1,$$

where

$$(59)\qquad P = \partial^2 g(z,\zeta)/\partial\xi\partial n + \sum_{\nu=1}^{n-1} \lambda_\nu(\partial\omega_\nu(z)/\partial n), \qquad z \in \Gamma,$$

and $\lambda, \cdots, \lambda_{n-1}$, is the minimizing set. We define the harmonic function $U(z)$ by the boundary values

$$(60)\qquad U(z) = a_\mu - b_\mu(|\,P\,|/P), \qquad z \in \Gamma, \qquad P \neq 0,$$

and the requirement that $|\,U(z)\,| \leq \operatorname{Max} |\,a_\mu\,| + b_\mu$ $(\mu = 1, \cdots, n)$ in D (This requirement is necessary since the definition (60) of the boundary values of $U(z)$ breaks down at the $2n$ possible zeros of P). Rewriting (58) in terms of this function $U(z)$, we obtain

$$\sum_{\mu=1}^{n} \tfrac{1}{2\pi} \int_{\Gamma_\mu} [a_\mu - U(z)]\,(\partial\omega_\nu/\partial n)\,ds - \sum_{\mu=1}^{n} (a_\mu/2\pi) \int_{\Gamma_\mu} (\partial\omega_\mu/\partial n)\,ds = 0,$$

whence $\int_{\Gamma} U(z)\,(\partial\omega_\nu/\partial n)\,ds = 0$, $\nu = 1, \cdots, n-1$, which shows that the harmonic conjugate $V(z)$ of $U(z)$ is single-valued in D. The analytic function $\Phi(z) = U(z) + iV(z)$ is thus single-valued in D; as shown by (60), it has furthermore the property (55), except at the points at which $P = 0$. The extremal property (54) follows by considering, in this order, formulas (57), (59), (60), (56).

It remains to be shown that no value is taken by $\Phi(z)$ more than n times in D. To this end, we observe from (60) that

$$P[U(z)-a_\mu] \leq 0, \qquad z \varepsilon \Gamma_\mu, \tag{61}$$

where P is defined in (59). As already pointed out on a previous occasion, a function of the type (59) cannot have more than $2n$ changes of sign on Γ. (61) and (55) show that each change of sign of P corresponds to a change of the value of $U(z)-a_\mu$ from b_μ to $-b_\mu$, or vice versa. Since two such changes correspond to one sheet of the conformal map of D given by $\Phi(z)$, it follows that the maximum number of sheets cannot exceed n, whence the assertion.

Incidentally, it is not without interest to point out that our procedure furnishes an existence proof for conformal mappings with the property (55).

The next problem concerns functions $\phi(z)$ which have a simple pole at a specified point of D and satisfy boundedness conditions of the type (53) on the boundary of D. For brevity, we take $a_\mu = 0$, $b_\mu = 1$. In order to obtain the corresponding result for more general values of these parameters, some features of the proof of Theorem VII have to be added.

THEOREM VIII. Let

$$\phi(z) = a/(z-\zeta) + O(|z-\zeta|) \qquad (\zeta \varepsilon D) \tag{62}$$

be single-valued and, except for the simple pole at $z=\zeta$, *regular in* D, *and let*

$$\limsup_{z\to\Gamma} |\operatorname{Re}\{\phi(z)\}| \leq 1. \tag{63}$$

Then

$$\operatorname{Re}\{a\} \leq \operatorname{Re}\{A\}, \tag{64}$$

where A *is the residue, at* $z=\zeta$, *of a function*

$$\Phi(z) = A/(z-\zeta) + O(|z-\zeta|) \tag{65}$$

which is single-valued and, except at $z=\zeta$, *regular in* D, *and satisfies*

$$|\operatorname{Re}\{\Phi(z)\}| = 1, \qquad z \varepsilon \Gamma, \tag{66}$$

with the possible exception of not more than $2n$ *points at which* $R\{\Phi(z)\}$ *may not be defined. The conformal map yielded by* $w=\Phi(z)$ *covers each of the half-planes* $\operatorname{Re}\{w\} > 1$ *and* $\operatorname{Re}\{w\} < 1$ *once, and it covers the strip* $-1 < \operatorname{Re}\{w\} < 1$ *not more than* n *times.*

Proof. The remark made at the beginning of the proof of Theorem VII applies also in this case, and we may apply Green's formula to the boundary values of $u(z) = \mathrm{Re}\{\phi(z)\}$. Using (16) and the residue theorem in order to shorten the computations, we obtain

$$-\frac{1}{2\pi}\int_{\Gamma} u(z)[\alpha(\partial g(z,\zeta)/\partial n) - \sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu(z)/\partial n)]\,ds$$

$$= -\frac{1}{2\pi i}\int_{\Gamma} u(z)\,[\alpha p'(z,\zeta) - \sum_{\nu=1}^{n-1}\lambda_\nu w'_\nu(z)]\,dz$$

$$= -\mathrm{Re}\{\frac{1}{2\pi i}\int_{\Gamma}\phi(z)\,[\alpha p'(z,\zeta) - \sum_{\nu=1}^{n-1}\lambda_\nu w'_\nu(z)\,dz\}$$

$$= \mathrm{Re}\{a[\alpha c + \sum_{\nu=1}^{n-1}\lambda_\nu w'_\nu(\zeta)]\},$$

where $\alpha, \lambda_1, \cdots, \lambda_{n-1}$, are arbitrary real parameters and

$$c = -\lim_{z\to\zeta} p'(z,\zeta) - 1/(z-\zeta). \tag{67}$$

If the arbitrary parameters are made subject to the condition

$$\alpha c + \sum_{\nu=1}^{n-1}\lambda_\nu w'_\nu(\zeta) = 1, \tag{68}$$

the above identity takes the form

$$\mathrm{Re}\{a\} = -\frac{1}{2\pi}\int_{\Gamma} u(z)\,[\alpha(\partial g(z,\zeta)/\partial n) - \sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu/\partial n)]\,ds. \tag{69}$$

In view of (63), it follows that

$$\mathrm{Re}\{a\} \leq \frac{1}{2\pi}\int_{\Gamma} |\,\alpha(\partial g(z,\zeta)/\partial n) - \sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu/\partial n)\,|\,ds. \tag{70}$$

Our next step is to minimize the right-hand side of (70) under the conditions (68). A formal computation shows that the necessary conditions for a minimum are

$$\frac{1}{2\pi}\int_{\Gamma} |\,P\,|/P\,(\partial\omega_\nu/\partial n)\,ds + \mathrm{Re}\{Aw'_\nu(\zeta)\} = 0, \qquad \nu = 1, \cdots, n-1 \tag{71}$$

$$\frac{1}{2\pi}\int_{\Gamma} |\,P\,|/P\,(\partial g(z,\zeta)/\partial n)\,ds + \mathrm{Re}\{Ac\} = 0, \tag{72}$$

where

$$P = \alpha(\partial g(z,\zeta)/\partial n) - \sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu(z)/\partial n), \qquad z \in \Gamma, \tag{72'}$$

and A is a complex Lagrange multiplier

We now define the bounded harmonic function $U_1(z)$ by the boundary values

$$U_1(z) = -|P|/P, \qquad z \in \Gamma, \qquad P \neq 0. \tag{73}$$

If $V_1(z)$ denotes the harmonc conjugate of $U_1(z)$, (73) and (5′) show that (71) and (72) can be rewritten in the forms

$$\frac{1}{2\pi} \int_{\Gamma_\nu} dV_1 = \operatorname{Re}\{Aw'_\nu(\zeta)\} \tag{74}$$

and

$$-\frac{1}{2\pi} \int_{\Gamma} U_1(z)\,(\partial g(z,\zeta)/\partial n)\,ds = \operatorname{Re}\{Ac\}, \tag{75}$$

respectively.

Consider now the harmonic functions

$$\psi_1(z) = \partial g(z,\zeta)/\partial\xi, \qquad \psi_2 = -\partial g(z,\zeta)/\partial\eta, \qquad \zeta = \xi + i\eta. \tag{76}$$

Both $\psi_1(z)$ and $\psi_2(z)$ are harmonic in D, except at $z = \zeta$ where $\psi_1(z) - \operatorname{Re}\{(z-\zeta)^{-1}\}$ and $\psi_2(z) - \operatorname{Im}\{(z-\zeta)^{-1}\}$ are harmonic. Since $g(z,\zeta)$ vanishes for $z \in \Gamma$ and any $\zeta \in D$, both $\psi_1(z)$ and $\psi_2(z)$ have the boundary values 0 at all points of Γ. The harmonic conjugates $\psi^*_1(z)$ and $\psi^*_2(z)$ of $\psi_1(z)$ and $\psi_2(z)$, respectively, are generally not single-valued. By the Cauchy-Riemann equations and the boundary properties of $\omega_\nu(z)$, we have

$$\int_{\Gamma_\nu} d\psi^*_1 = \int_{\Gamma_\nu} (\partial\psi^*_1/\partial s)\,ds = \int_{\Gamma_\nu} (\partial\psi_1/\partial n)\,ds = \int_{\Gamma} \omega_\nu(z)\,(\partial\psi_1/\partial n)\,ds.$$

Similarly, $\int_{\Gamma_\nu} d\psi^*_2 = \int_{\Gamma} \omega_\nu(z)\,(\partial\psi_2/\partial n)\,ds$. Hence, by (76),

$$\int_{\Gamma_\nu} d\psi^*_1 = \partial/\partial\xi\Big[\int_{\Gamma} \omega_\nu(z)\,(\partial g(z,\zeta)/\partial n)\,ds\Big],$$

$$\int_{\Gamma_\nu} d\psi^*_2 = -\partial/\partial\eta\Big[\int_{\Gamma} \omega_\nu(z)\,(\partial g(z,\zeta)/\partial n)\,ds\Big],$$

whence, by Green's formula,

$$\frac{1}{2\pi} \int_{\Gamma_\nu} d\psi^*_1 = -\partial\omega_\nu(\zeta)/\partial\xi = -\operatorname{Re}\{w'_\nu(\zeta)\},$$

$$\frac{1}{2\pi} \int_{\Gamma_\nu} d\psi^*_2 = \partial\omega_\nu(\zeta)/\partial\eta = -\operatorname{Im}\{w'_\nu(\zeta)\}.$$

If $A = A_1 + iA_2$, it follows from these formulas and from (74) that

$$\int_{\Gamma_\nu} d(V_1 + A_1 d\psi^*_1 - \mathrm{A}_2 d\psi^*_1) = 0, \qquad \nu = 1, \cdots, n-1. \tag{77}$$

We now introduce the harmonic function $U(z) = U_1(z) + A_1\psi_1(z) - A_2\psi_2(z)$, its harmonic conjugate $V(z) = V_1(z) + A_1\psi^*_1(z) - A_2\psi^*_2(z)$, and the corresponding analytic function $\Phi(z) = U(z) + iV(z)$. Since

$$A_1\psi_1(z) - A_2\psi_2(z) = A_1 \operatorname{Re}\{1/(z-\zeta)\} - A_2 \operatorname{Im}\{1/(z-\zeta)\} + \cdots$$

$= \operatorname{Re}\{A/(z-\zeta)\} + c_1 + O(|z-\zeta|)$, the function $\Phi(z)$ is of the form

$$\Phi(z) = A/(z-\zeta) + c_1 + O(|z-\zeta|). \tag{78}$$

Since both $\psi_1(z)$ and $\psi_2(z)$ vanish on Γ, $U(z)$ has the same boundary values as $U_1(z)$; hence, by (73),

$$U(z) = -|P|/P, \qquad z \in \Gamma, \qquad P \neq 0. \tag{79}$$

(77) is equivalent to $\int_{\Gamma_\nu} dV = 0$, $\nu = 1, \cdots, n-1$, which shows that $V(z)$, and therefore also the analytic function $\Phi(z)$, is single-valued in D. The condition (75) takes the form

$$-\tfrac{1}{2\pi}\int_\Gamma U(z)(\partial g(z,\zeta)/\partial n)\,ds = \operatorname{Re}\{Ac\},$$

which, in view of

$$-\tfrac{1}{2\pi}\int_\Gamma U(z)(\partial g(z,\zeta)/\partial n)\,ds = -\operatorname{Re}\{\tfrac{1}{2\pi i}\int_\Gamma \Phi(z)p'(z,\zeta)\,dz\}$$

$= \operatorname{Re}\{Ac + c_1\}$, shows that the real part of the constant c_1 in (78) vanishes. Since, so far, the function $\Phi(z)$ was only determined up to an arbitrary constant, we may choose this constant so as to make the imaginary part of c_1 likewise vanish. We have thus $c_1 = 0$, which shows that $\Phi(z)$ is indeed of the form (65).

(79) shows that $\Phi(z)$ has the property (66). The extremal property (64) follows by considering, in that order, formulas (70), (72′), (79), (78), (68), (69), and observing that, in (78), $c_1 = 0$. As regards the number of coverings by the conformal map of D as yielded by $w = \Phi(z)$, it is clear from (66) and the fact that $\Phi(z)$ has exactly one simple pole in D, that all points outside the strip $-1 \leq \operatorname{Re}\{w\} \leq 1$ are covered exactly once. We further have, by (73) and (72′),

$$U(z)\left[\alpha(\partial g(z,\zeta)/\partial n) - \sum_{\nu=1}^{n-1}\lambda_\nu(\partial\omega_\nu(z)/\partial n)\right] \leq 0, \qquad z \in \Gamma. \tag{80}$$

It is well-known, and easily proved, that the expression in the square bracket cannot change its sign on Γ more than $2(n-1)$ times. It follows from (80) that this is also the maximum number of changes of sign of $U(z)$ on Γ. Each two of such changes correspond to one sheet of the conformal map of $\Phi(z)$ which covers the strip $-1 < \operatorname{Re}\{w\} < 1$. The simple pole of $\Phi(z)$ contributes one extra covering of the strip by one sheet of the map which contains the point at infinity. It follows that the total number of coverings of the strip cannot exceed n. This completes the proof of Theorem VIII.

WASHINGTON UNIVERSITY.

REFERENCES.

[1] L. V. Ahlfors, "Bounded analytic functions," *Duke Mathematical Journal*, vol. 14 (1947), pp. 1-11.

[2] P. Cacridis-Theodorakopulos, "Über die untere Grenze der Rundungsschranke der beschränkten Funktionen $f(z)$, deren $|f'(0)|$ vorgegeben ist," *Mathematische Annalen*, vol. 113 (1937), pp. 657-664.

[3] J. Dieudonné, "Recherches sur quelques problèmes relatifs aux polynomes et aux fonctions bornées," *Ann. le l'École Normale supérieure*, ser. 3, vol. 48 (1931), pp. 247-358.

[4] P. R. Garabedian, "The Schwarz lemma and the Szegö kernel function," *Transactions of the American Mathematical Society*, vol. 67 (1949), pp. 1-35.

[5] H. Grunsky, "Eindeutige beschränkte Funktionen in mehrfach zusammenhängenden Gebieten. I," *Jahresbericht der Deutschen Math. Ver.*, vol. 50 (1940), pp. 230-255.

[6] H. Grunsky, "Eindeutige beschränkte Funktionen in mehrfach zusammenhängenden Gebieten. II," *ibid.*, vol. 52 (1942), pp. 118-132.

[7] E. Landau, "Der Picard-Schottkysche Satz und die Blochsche Konstante," *Sitzungsber. Akad. d. Wiss. Berlin, Phys.-Math. Kl.* (1926), pp. 467-474.

[8] Z. Nehari, "Analytic functions possessing a positive real part," *Duke Mathematical Journal*, vol. 15 (1948), pp. 1033-1042.

[9] R. Nevanlinna, "Über beschränkte Funktionen, die in gegebenen Punkten vorgeschriebene Werte annehmen," *Ann. Acad. Sci. Fennicae*, vol. 13 (1919), No. 1.

[10] G. Pick, "Über die Beschränkungen analytischer Funktionen, welche durch vorgegebene Funktionswerte bewirkt werden," *Mathematische Annalen*, vol. 77 (1916), pp. 7-23.

ASYMPTOTIC IDENTITIES AMONG PERIODS OF INTEGRALS OF THE FIRST KIND.*

By P. R. Garabedian.

1. Introduction.

With each closed Riemann surface S of genus g there are associated g normal integrals of the first kind, $w_1(z), w_2(z), \cdots, w_g(z)$. The integrals $w_\nu(z)$ have the periods $\delta_{\mu\nu}\pi i$ along one set of g canonical cuts α_μ on S, while along the remaining g canonical cuts β_μ, which intersect the α_μ, these integrals have the periods $a_{\mu\nu}$. It is well known that $a_{\mu\nu} = a_{\nu\mu}$ and that the real quadratic form

$$\sum_{\mu,\nu=1}^{g} x_\mu x_\nu \operatorname{Re}\{a_{\mu\nu}\}$$

is positive definite.

For $g > 1$, the closed Riemann surface S depends in its conformal type upon $3g - 3$ complex parameters, while there are altogether $g(g+1)/2$ different quantities $a_{\mu\nu}$. Thus it is a classical problem to determine the $\frac{1}{2}g(g+1) - (3g-3) = \frac{1}{2}(g-2)(g-3)$ identities which must hold among the periods $a_{\mu\nu}$. Only in one case is this problem solved, namely, in the case $g = 4$ of surfaces of lowest genus for which an identity can be expected. Here Schottky [6] has found the one necessary identity in terms of the Riemann theta-functions associated with S.

Aside from the complete treatment by Schottky of the case $g = 4$, there is a paper [5] of Poincaré in which an identity among the $a_{\mu\nu}$ is found for the asymptotic situation where the diagonal terms $a_{\nu\nu}$ remain finite (different from zero), while the $a_{\mu\nu}$ with $\mu \neq \nu$ tend towards zero. Poincaré's identity involves six of the $a_{\mu\nu}$ with $\mu \neq \nu$ and has the form

$$(a_{12}a_{13}a_{24}a_{34})^{\frac{1}{2}} + (a_{13}a_{14}a_{24}a_{23})^{\frac{1}{2}} = (a_{12}a_{14}a_{23}a_{34})^{\frac{1}{2}}, \tag{1}$$

where we neglect terms of higher order.

With each plane domain D of finite connectivity n bounded by proper continua $C_1, \cdots, C_n$ we can associate in a unique manner a symmetric closed Riemann surface S of genus $g = n - 1$. The surface S can be obtained, for example, as follows. We map D conformally and $(n, 1)$ upon

---

* Received March 25, 1950.

the interior of the unit circle so that each continuum C_μ corresponds to the unit circle traversed once. We then invert the image of D in the unit circle, and we identify corresponding pairs of boundary points of the two domains so obtained. We thus construct a model with $n = g + 1$ sheets of the required closed Riemann surface S.

If $g = n - 1$ of the curves C_μ on S are chosen to be the canonical cuts β_μ on S, then the harmonic functions

$$\omega_\mu(z) = \operatorname{Re}\{-2w_\mu(z)i/\pi\}$$

are characterized (modulo 2) by the property that they have boundary values $\delta_{\mu\nu}$ on the C_ν and are single-valued in D. Hence the periods

$$p_{\mu\nu} = (1/2\pi) \oint_{C_\nu} (\partial\omega_\mu/\partial n)\, ds$$

of the harmonic conjugates of the functions $\omega_\mu(z)$ are related to the quantities $a_{\mu\nu}$ according to the rule $p_{\mu\nu} = a_{\mu\nu}/\pi^2$. Thus a special case of the identities among the $a_{\mu\nu}$ will be a set of identities among the real quantities $p_{\mu\nu}$.

In the present paper we shall develop asymptotic identities among the $p_{\mu\nu}$ for various situations in which the plane domain D degenerates. As a special case, we shall arrive at Poincaré's identity among the $p_{\mu\nu}$ when suitable assumptions are made upon the limiting position of the domain D. Our aim throughout will be to obtain an insight into the nature of the identities among the $a_{\mu\nu}$ in general, and we shall be able to avoid all restriction upon the connectivity n of the domains we consider.

Our analysis will be, on the whole, altogether elementary. We shall emphasize the simple geometric interpretation of our asymptotic relations and their connections with conformal mapping. We remark that an important aspect of our problem, independent of the above discussion of closed Riemann surfaces, arises from the attempt at interpretation of the invariants $p_{\mu\nu}$ as moduli for the conformal type of D [1, 2, 3].

2. Domains with infinitesimal boundaries. We suppose that a domain D is given which is bounded by the n proper continua $C_1, \cdots, C_n$. There will be no loss of generality if we assume that the C_ν are circles with, say, centers a_ν, and that $a_n = \infty$, for any domain bounded by n proper continua can be mapped conformally upon such a domain. We denote the radii of the circles C_ν, $\nu < n$, by $t\epsilon_\nu$, while we take the radius of C_n to be $R = 1/(t\epsilon_n)$, and we attempt to express the invariants $p_{\mu\nu}$ asymptotically in terms of t as $t \to 0$.

For each $\nu < n$, the harmonic measures $\omega_\nu(z)$ in D have the form

$$\omega_\nu(z) = \sum_{\mu=1}^{n-1} \lambda_{\mu\nu}(\log | R^2 - \bar{a}_\mu z | - \log | R(z - a_\mu) |) + O(t),$$

for suitable constants $\lambda_{\mu\nu}$. Since $R = 1/(t\epsilon_n)$, we have therefore

$$(2) \qquad \omega_\nu(z) = -\sum_{\mu=1}^{n-1} \lambda_{\mu\nu} \log | z - a_\mu | + \lambda_{n\nu} \log t\epsilon_n + O(t),$$

where $\sum_{\mu=1}^{n-1} \lambda_{\mu\nu} = 0$. The $\lambda_{\mu\nu}$ are determined by the conditions that $\omega_\nu(z)$ vanishes for $z \varepsilon C_\mu$, $\mu \neq \nu$, and has the value 1 on C_ν. These conditions yield the system of linear equations

$$(3) \qquad \sum_{\mu\neq k,\mu=1}^{n-1} \lambda_{\mu\nu} \log | a_\mu - a_k | + \lambda_{k\nu} \log t\epsilon_k - \lambda_{n\nu} \log t\epsilon_n = \delta_{k\nu},$$
$$k = 1, \cdots, n-1,$$

since we neglect terms $O(t)$.

From (2) we obtain for the periods $p_{\mu\nu}$ the asymptotic formulas

$$p_{\mu\nu} = \lambda_{\mu\nu} + O(t),$$

while from (3) the $\lambda_{\mu\nu}$ must have the form

$$\lambda_{\mu\nu} = \sigma_{\mu\nu}/| \log t | + \rho_{\mu\nu}/| \log t |^2 + O(1/| \log t |^3).$$

Considering the terms in (3) of the lowest order, namely, of the order 0 in $1/| \log t |$, we have $\sigma_{k\nu} - \sigma_{n\nu} = \delta_{k\nu}$, $k = 1, \cdots, n-1$, while at the same time we require $\sum_{\mu=1}^{n} \sigma_{\mu\nu} = 0$. Hence $\sigma_{\mu\nu} = \delta_{\mu\nu} - n^{-1}$, $\mu, \nu = 1, \cdots, n$. Thus the first order terms for the $p_{\mu\nu}$ throw no light upon the identities among these periods, and we must investigate the terms of second order.

From the terms of order 1 in $1/| \log t |$ in (3) we find for the asymptotic terms

$$(4) \qquad \rho_{\mu\nu} = \lim_{t\to 0}\{p_{\mu\nu} - (\delta_{\mu\nu} - n^{-1})(1/| \log t |)\}| \log t |^2$$

the linear equations

$$\sum_{\mu\neq k,\mu=1}^{n-1} \sigma_{\mu\nu} \log | a_\mu - a_k | + \sigma_{k\nu} \log \epsilon_k - \sigma_{n\nu} \log \epsilon_n - \rho_{k\nu} + \rho_{n\nu} = 0,$$

$k, \nu = 1, \cdots, n-1$, and we require further, of course, $\sum_{\mu=1}^{n} \rho_{\mu\nu} = 0$. We sum the first equations on k and combine with the second equation to find

$$\rho_{n\nu} = -n^{-1} \sum_{k,\mu=1,\mu\neq k}^{n-1} \sigma_{\mu\nu} \log | a_\mu - a_k | - n^{-1} \sum_{k=1}^{n-1} \sigma_{k\nu} \log \epsilon_k - n^{-2}(n-1)\log \epsilon_n$$

$$= n^{-2} \sum_{k,\mu=1,\mu\neq k}^{n-1} \log | a_\mu - a_k | - n^{-1} \sum_{k=1,k\neq\nu}^{n-1} \log | a_k - a_\nu | + n^{-2} \sum_{k=1}^{n-1} \log \epsilon_k$$

$$- n^{-1} \log \epsilon_\nu - n^{-2}(n-1)\log \epsilon_n.$$

Hence

$$\rho_{\mu\nu} = n^{-2} \sum_{j,k=1,j\neq k}^{n-1} \log | a_j - a_k | + n^{-2} \sum_{k=1}^{n} \log \epsilon_k - n^{-1} \sum_{k=1,k\neq\nu}^{n-1} \log | a_k - a_\nu |$$

$$- n^{-1} \sum_{k=1,k\neq\mu}^{n-1} \log | a_k - a_\mu | - (\delta_{\mu\nu} - 1)\log | a_\mu - a_\nu | - n^{-1} \log \epsilon_\nu$$

$$- n^{-1} \log \epsilon_\mu + \delta_{\mu\nu} \log \epsilon_\mu,$$

or, setting

$$\alpha_{\mu\nu} = - n^{-1} \log | a_\mu - a_\nu |, \quad \mu \neq \nu; \qquad \alpha_{\mu\mu} = - n^{-1} \log \epsilon_\mu;$$

$$A = n^{-2} \sum_{\mu,\nu=1,\mu\neq\nu}^{n-1} \log | a_\mu - a_\nu | + n^{-2} \sum_{\mu=1}^{n} \log \epsilon_\mu,$$

we obtain

$$(5) \qquad \rho_{\mu\nu} = A + \sum_{k=1}^{n-1} \alpha_{k\mu} + \sum_{k=1}^{n-1} \alpha_{k\nu} - n\alpha_{u\nu}.$$

In particular, $\rho_{\mu\nu} = \rho_{\nu\mu}$, as was to be expected.

Our object now will be to solve the equations (5) for the $\alpha_{\mu\nu}$ with $\mu \neq \nu$, to obtain identities among these purely geometrical quantities, and to express the identities, finally, in terms of the asymptotic invariants $\rho_{\mu\nu}$.

We set $\alpha_\mu = \sum_{\nu=1}^{n-1} \alpha_{\mu\nu}$. From (5) we have, summing on ν from 1 to $n-1$,

$$-\rho_{\mu n} = (n-1)A + (n-1)\alpha_\mu + \sum_{\mu,\nu=1}^{n-1} \alpha_{\mu\nu} - n\alpha_\mu$$

$= (n-1)A + \sum_{\mu,\nu=1}^{n-1} \alpha_{\mu\nu} - \alpha_\mu$. Substituting the value of α_μ so obtained back into (5), we find

$$n\alpha_{\mu\nu} = A + \alpha_\mu + \alpha_\nu - \rho_{\mu\nu} = A + 2(n-1)A + 2\sum_{\mu,\nu=1}^{n-1} a_{\mu\nu} + \rho_{\mu n} + \rho_{\nu n} - \rho_{\mu\nu}.$$

Noting that $\sum_{\mu,\nu=1}^{n-1} \alpha_{\mu\nu} = -nA - \alpha_{nn}$ and introducing the notation $B = A + 2\alpha_{nn}$, we obtain

$$(6) \qquad \log | a_\mu - a_\nu | = \rho_{\mu\nu} + \sum_{k=1}^{n-1} \rho_{k\mu} + \sum_{k=1}^{n-1} \rho_{k\nu} + B,$$

whence $d_{\mu\nu} = | a_\mu - a_\nu |^2 = \exp[2(\rho_{\mu\nu} + \sum_{k=1}^{n-1} \rho_{k\mu} + \sum_{k=1}^{n-1} \rho_{k\nu} + B)]$.

Now it is not difficult to derive polynomial identities among the squares $d_{\mu\nu}$ of the distances between the points $a_1, \cdots, a_{n-1}$. Indeed, since the volume of the degenerate tetrahedron whose vertices are any four of the points a_ν, for example, whose vertices are a_1, a_2, a_3 and a_4, is zero, we have the five-by-five determinant relation

$$\begin{vmatrix} -2\operatorname{Re}(a_\mu \bar{a}_\nu) & 1 \\ 1 & 0 \end{vmatrix} = 0, \quad \mu, \nu = 1, \cdots, 4.$$

Here the term in the upper left-hand corner of the determinant represents the four-by-four matrix of quantities $-2\operatorname{Re}(a_\mu \bar{a}_\nu)$, and the fifth row and fifth column are formed of units, except for the vanishing diagonal term. Multiplying the fifth row of this determinant by $|a_\mu|^2$ and adding to the μ-th row, and multiplying the fifth column by $|a_\nu|^2$ and adding to the ν-th column, we obtain

$$(7) \qquad \begin{vmatrix} |a_\mu - a_\nu|^2 & 1 \\ 1 & 0 \end{vmatrix} = 0, \text{ i. e.,} \quad \begin{vmatrix} d_{\mu\nu} & 1 \\ 1 & 0 \end{vmatrix} = 0, \qquad \mu, \nu = 1, \cdots, 4,$$

where $d_{\mu\mu} = 0$. Thus there is a polynomial (7) in six variables and of degree three which vanishes upon substitution of values of the six distances amongst any four of the $n - 1$ points $a_1, \cdots, a_{n-1}$. Alternately, we can express this fact by saying that the n-by-n matrix

$$\left\| \begin{matrix} d_{\mu\nu} & 1 \\ 1 & 0 \end{matrix} \right\|, \qquad \mu, \nu = 1, \cdots, n-1,$$

has rank 4.

Altogether, we obtain in this way $\frac{1}{2}(n-4)(n-3) = \frac{1}{2}(g-3)(g-2)$ independent polynomial identities among the quantities $d_{\mu\nu}$. Using (6), we can now obtain from (7) a like number of identities among the limiting invariants $\rho_{\mu\nu}$. Dividing (7) by e^{6B} we find

$$(8) \qquad \begin{vmatrix} \gamma_{\mu\nu} \exp[2(\rho_{\mu\nu} + \sum_{k=1}^{n-1} (\rho_{k\mu} + \rho_{k\nu}))] & 1 \\ 1 & 0 \end{vmatrix} = 0,$$

$$\gamma_{\mu\nu} = 1 - \delta_{\mu\nu} = \begin{cases} 0, & \mu = \nu \\ 1, & \mu \neq \nu, \end{cases}$$

where μ, ν can run over any four distinct integers from 1 to $n - 1$. These identities can also be expressed in the form

$$(9)\quad \sum_{6} \exp[2(\rho_{ij} + 2\rho_{kl} - \sum_{\mu=1}^{n-1} (\rho_{\mu i} + \rho_{\mu j}))] + \sum_{4} \exp[2(\rho_{ij} + \rho_{jk} + \rho_{ki} - 2\sum_{\mu=1}^{n-1} \rho_{\mu l})] = \sum_{12} \exp[2(\rho_{ij} + \rho_{jk} + \rho_{kl} - \sum_{\mu=1}^{n-1} (\rho_{\mu i} + \rho_{\mu l}))],$$

where the subscripts under the summation signs indicate how many terms occur, and where the indices i, j, k, l are always distinct and range over any one set of four integers between 1 and $n-1$. Finally, we can use the equivalent expressions

$$(10)\quad \sum_{10} \exp[2(2\rho_{ij} + \rho_{kl} + \rho_{lm} + \rho_{mk})] = \sum_{12} \exp[2(\rho_{ij} + \rho_{jk} + \rho_{kl} + \rho_{lm} + \rho_{mi})],$$

where i, j, k, l and m are distinct and range over any one set of five integers between 1 and n. Formulas (10) and (9) are the same when $m = n$, and (10) follows from (9) when we remark that the choice of C_n as the circle centered at infinity is arbitrary. These identities can be summarized in the statement that the n-by-n matrix $\| \gamma_{\mu\nu} \exp(2\rho_{\mu\nu}) \|$ has rank 4.

Any of the equivalent sets of identities (8), (9) or (10) represent asymptotic relations into which the identities mentioned in Section 1 must degenerate as $t \to 0$. From another point of view, one can interpret the linear equations (3) as a parametric form of these identities which would include all the terms in $1/|\log t|$ in the asymptotic expansions of the $p_{\mu\nu}$ in terms of t. However, the identities (10) would appear to yield more insight into the nature of the finite identities among the $p_{\mu\nu}$, especially so in view of the occurrence of exponentials of the $\rho_{\mu\nu}$ and the consequent closeness to the known result obtained by Schottky in terms of Riemann theta-functions.

3. Domains in the unit circle. We proceed to discuss a new case in which the n-th boundary curve C_n of D is the unit circle $|z| = 1$, while the remaining C_ν are circles of radius $t\epsilon_\nu$ about points a_ν in $|z| < 1$, $\nu = 1, \cdots, n-1$. In such domains we have for the harmonic measures $\omega_\nu(z)$ with $\nu < n$

$$\omega_\nu(z) = \sum_{k=1}^{n-1} \lambda_{k\nu}(\log|1 - \bar{a}_k z| - \log|z - a_k|) + O(t),$$

where

$$(11)\quad \sum_{k=1, k\neq\mu}^{n-1} \lambda_{k\nu}(\log|1 - \bar{a}_k a_\mu| - \log|a_\mu - a_k|) + \lambda_{\mu\nu}(\log(1 - |a_\mu|^2) - \log t\epsilon_\mu) = \delta_{\mu\nu}, \qquad \mu = 1, \cdots, n-1.$$

Again $p_{\mu\nu} = \lambda_{\mu\nu} + O(t)$, $\mu, \nu = 1, \cdots, n-1$, and we have the asymptotic developments

$$\lambda_{\mu\nu} = \delta_{\mu\nu}/|\log t| + P_{\mu\nu}/|\log t|^2 + O(1/|\log t|^3).$$

From the terms in $1/|\log t|$ in (11) we obtain the linear equations

$$\sum_{k=1, k\neq\mu}^{n-1} \delta_{k\nu}(\log|1-\bar{a}_k a_\mu| - \log|a_\mu - a_k|) + \delta_{\mu\nu}(\log(1-|a_\mu|^2) - \log \epsilon_\mu) + P_{\mu\nu} = 0,$$

$\mu, \nu = 1, \cdots, n-1$. Hence for $\mu = \nu$ we get

$$P_{\mu\mu} = -\log(1-|a_\mu|^2) + \log \epsilon_\mu;$$

$$P_{\mu\nu} = \log|a_\nu - a_\mu| - \log|1 - \bar{a}_\nu a_\mu|, \qquad \mu \neq \nu.$$

Thus, in particular, $P_{\mu\nu} = p_{\mu\nu}/(p_{\mu\mu}p_{\nu\nu}) + O(1/|\log t|)$, $\mu \neq \nu$, and $P_{\mu\nu} = P_{\nu\mu}$.

Note that the $P_{\mu\mu}$ are independent of the $P_{\mu\nu}$ $(\mu \neq \nu)$, since ϵ_μ can be chosen arbitrarily. Hence identities in this case will involve only the non-diagonal terms $P_{\mu\nu}$. We shall therefore assume from this point on in writing $P_{\mu\nu}$ that $\mu \neq \nu$.

We set $c_{\mu\nu} = |(a_\mu - a_\nu)/(1-\bar{a}_\mu a_\nu)|^2 = \exp(2P_{\mu\nu})$, and we note that $c_{\mu\nu}$ is closely connected with the non-Euclidean metric in the unit circle. Indeed, $c_{\mu\nu}$ is invariant when a_μ and a_ν are subjected to a conformal transformation of the interior of the unit circle upon itself. The negative of the invariant $P_{\mu\nu}$ is nothing more than the Green's function $G(a_\mu, a_\nu)$ in the unit circle.

We determine now identities amongst the squared conformal distances $c_{\mu\nu}$. There is no loss of generality if we assume that one point $a_k = 0$, $k < n$, in view of the above remarks. We thus obtain, in the notation $d_{\mu\nu} = |a_\mu - a_\nu|^2$ of Section 2, the relations

$$\begin{aligned}(1/c_{\mu\nu}) - 1 &\\ &= (1 - 2\operatorname{Re}(a_\mu \bar{a}_\nu) + |a_\mu|^2|a_\nu|^2)/(|a_\mu|^2 - 2\operatorname{Re}(a_\mu\bar{a}_\nu) + |a_\nu|^2) - 1\\ &= (1-d_{k\mu})(1-d_{k\nu}/d_{\mu\nu}, \text{ or } d_{\mu\nu} = c_{\mu\nu}(1-d_{k\mu})(1-d_{k\nu})/(1-c_{\mu\nu}).\end{aligned}$$

We substitute this expression in the five-by-five determinant of the identity

$$\begin{vmatrix} d_{\mu\nu} & 1 \\ 1 & 0 \end{vmatrix} = 0$$

of Section 2, we divide each of the first four rows by a suitable factor

$(1 - d_{k\mu})$ and each of the first four columns by a suitable factor $(1 - d_{k\nu})$, and we obtain

$$\begin{vmatrix} c_{\mu\nu}/(1 - c_{\mu\nu}) & 1/(1 - d_{k\mu}) \\ 1/(1 - d_{k\nu}) & 0 \end{vmatrix} = 0.$$

But $d_{k\nu} = c_{k\nu}$, $d_{k\mu} = c_{k\nu}$, since $a_k = 0$. If the fourth row and fourth column of the last determinant correspond to the index k, we have

$$\begin{vmatrix} c_{\mu\nu}/(1 - c_{\mu\nu}) & c_{\mu k}/(1 - c_{\mu k}) & 1/(1 - c_{\mu k}) \\ c_{\nu k}/(1 - c_{\nu k}) & 0 & 1 \\ 1/(1 - c_{\nu k}) & 1 & 0 \end{vmatrix} = 0,$$

where now the upper left-hand term stands for a three-by-three matrix, while the remaining terms in $c_{\nu k}$, $c_{\mu k}$ correspond to row- and column-vectors in 3-space. Substracting the fourth row from the fifth row and the fourth column from the fifth column, we obtain, finally,

$$(12) \qquad \begin{vmatrix} c_{\mu\nu}/(1 - c_{\mu\nu}) & 1 \\ 1 & -2 \end{vmatrix} = 0,$$

where, once more, the upper left-hand term represents a four-by-four matrix. Since k is arbitrary, there is no restriction whatever on this identity.

If we write $c_{\mu\nu}/(1 - c_{\mu\nu}) = 1/(c_{\mu\nu}^{-1} - 1)$, and note that $c_{\mu\nu} < 1$, we see that (12) is actually a polynomial identity of degree nine in six variables $1/c_{\mu\nu}$. We now substitute the values of the $c_{\mu\nu}$ in terms of the $P_{\mu\nu}$ into the purely geometrical identities (12) and obtain the relations

$$(13) \qquad \begin{vmatrix} \gamma_{\mu\nu}/(e^{-2P_{\mu\nu}} - 1) & 1 \\ 1 & -2 \end{vmatrix} = 0, \quad \gamma_{\mu\nu} = 1 - \delta_{\mu\nu},$$

which are, then, polynomial identities of degree nine, each involving six exponentials $e^{-2P_{\mu\nu}}$ with $\mu \neq \nu$. There are precisely $\frac{1}{2}(n - 3)(n - 4) = \frac{1}{2}(g - 2)(g - 3)$ independent identities (13), as is evident from our geometrical derivation.

Thus for the present limiting case, the identities among the $p_{\mu\nu}$ reduce to the relations (13). Indeed, we have

$$(14) \qquad \begin{vmatrix} \gamma_{\mu\nu}/(\exp[-2p_{\mu\nu}/(p_{\mu\mu}p_{\nu\nu})] - 1) & 1 \\ 1 & -2 \end{vmatrix} + O(1/|\log t|) = 0.$$

Of course, terms of higher order can be discussed starting from the parametric form (11) of the identities.

We note in passing that many further asymptotic situations of the domain D can be considered, and corresponding identities will result. For example, if one allows k of the continua C_ν bounding D to shrink to points, while $n - k$ remain fixed, one obtains parametric identities involving the Green's function of the domain bounded by the fixed continua. However, having outlined our method, we shall restrict ourselves here to but one more case, which is of particular interest because it leads to the identities of Poincaré [5].

4. Poincaré's identity. We consider a situation which is analogous to the requirement that the a_ν of Section 3 tend to $|z| = 1$ at the same rate at which the radii of the circles C_ν tend to 0. However, it is no longer convenient to use canonical domains bounded by circles in our analysis.

Let D lie in the upper half-planes $\operatorname{Im}(z) > 0$, let C_n be the real axis, and let the C_ν with $\nu < n$ be horizontal line segments in $\operatorname{Im}(z) > 0$. Every domain of connectivity n bounded by proper continua can be mapped conformally upon such a canonical region. We suppose that the segment C_ν has length $l_\nu t$ and is at a height $h_\nu t$ above the real axis. We shall study the periods $p_{\mu\nu}$ as $t \to 0$ and as the C_ν with $\nu \leqq n - 1$ shrink towards points b_ν on the real axis.

If C denotes a horizonal segment of length l at height one unit above the real axis, we denote by $\omega(z, l)$ that function which is harmonic in the half-plane $\operatorname{Im}(z) > 0$, except on C, and which assumes on C the value 1, while on the real axis it vanishes. We then find for the harmonic measures $\omega_\nu(z)$ in the domain D the asymptotic formulas

$$\omega_\nu(z) = \omega((z - b_\nu)/(h_\nu t), l_\nu/h_\nu) - \sum_{\mu=1, \mu\neq\nu}^{n-1} \omega((z - b_\mu)/(h_\mu t), l_\mu/h_\mu)\omega((b_\mu + h_\mu t i - b_\nu)/(h_\nu t), l_\nu/h_\nu) + O(t^3). \tag{15}$$

Thus, if we take $l = l_\nu/h_\nu$,

$$\lim_{t\to 0} p_{\nu\nu} = (1/2\pi) \oint_C [\partial\omega(z, l_\nu/h_\nu)/\partial n] ds = \pi_{\nu\nu}$$

is different from zero and depends only upon the ratio l_ν/h_ν. On the other hand, the $p_{\mu\nu}$ with $\mu \neq \nu$ are infinitesimal and we have

$$p_{\mu\nu} = - \pi_{\mu\mu}\, \omega((b_\mu - b_\nu)/h_\nu t) + (h_\mu/h_\nu)i, l_\nu/h_\nu) + O(t^3).$$

Let us denote by $W(z, l)$ the analytic function with

$$\operatorname{Re}\{W(z, l)\} = \omega(z, l), \qquad \lim_{z\to\infty} W(z, l) = 0.$$

We find, applying Schwarz's principle of reflection, that the function

$$F(z) = W(z,1) - p\log(z+i) + p\log(z-i)$$

is analytic and single-valued in the entire z-plane slit along the segment C and its reflection $\tilde{C}$ in the real axis. Here

$$p = (1/2\pi)\oint_C [\partial\omega(z,l)/\partial n]\,ds,$$

where integration is to be carried out along both edges of the slit C. Clearly $F(\infty) = 0$, and we shall proceed to show that

$$\eta_1 = \lim_{z\to\infty} zF(z) = 0, \qquad \eta_2 = \lim_{z\to\infty} z^2F(z) = 0.$$

Since $W(z,l)$ is pure imaginary for z real, we see that η_1 and η_2 are pure imaginary. Hence from the residue theorem we obtain

$$\begin{aligned}\eta_1 &= 1/(2\pi i)\oint_{C+\tilde{C}} \{W(z,l) - p\log(z+i) + p\log(z-i)\}\,dz \\ &= 1/(2\pi i)\oint_{C+\tilde{C}} \{\omega(z,l) - p\log|z+i| + p\log|z-i|\}\,dx,\end{aligned}$$

since $dy = 0$ along $C+\tilde{C}$. But since we integrate in both directions along C and $\tilde{C}$, and since $\omega(z,l)$ is 1 on C and -1 on $\tilde{C}$, we find that the integral vanishes. Thus, indeed, $\eta_1 = 0$. Furthermore, by symmetry $F(iy)$ is real, and hence, since η_2 is imaginary, it must vanish.

We now obtain, then,

$$\begin{aligned}&\omega((b_\mu - b_\nu)/(h_\nu t) + (h_\mu/h_\nu)i,\, l_\nu/h_\nu) \\ &\quad = \pi_{\nu\nu}(\log|(b_\mu - b_\nu)/(h_\nu t) + (h_\mu/h_\nu)i + i| \\ &\qquad\quad - \log|(b_\mu - b_\nu)/(h_\nu t) + (h_\mu/h_\nu)i - i|) + O(t^3) \\ &\quad = (\pi_{\nu\nu}/2)\log|1 + (2i)/((b_\mu - b_\nu)/(h_\nu t) + (h_\mu/h_\nu)i - i)|^2 + O(t^3)\end{aligned}$$

$= (2\pi_{\nu\nu}h_\mu h_\nu t^2)/(b_\mu - b_\nu)^2 + O(t^3)$. Thus

$$p_{\mu\nu} = -(2\pi_{\nu\nu}\pi_{\mu\mu}h_\nu h_\mu t^2)/(b_\mu - b_\nu)^2 + O(t^3). \tag{16}$$

We set

$$\pi_{\mu\nu} = \lim_{t\to 0} p_{\mu\nu}t^{-2} = -(2\pi_{\mu\mu}\pi_{\nu\nu}h_\mu h_\nu)/(b_\mu - b_\nu)^2. \tag{17}$$

Now if $b_i < b_j < b_k < b_l$, then clearly

$$(b_l - b_i)(b_k - b_j) + (b_j - b_i)(b_l - b_k) = (b_k - b_i)(b_l - b_j),$$

and thus, introducing the reciprocal squared distances $e_{ij} = 1/(b_i - b_j)^2$ and dividing by

$$(b_j - b_i)(b_k - b_j)(b_l - b_k)(b_k - b_i)(b_l - b_i)(b_l - b_j),$$

we obtain the geometrical relation

$$(18) \qquad (e_{ij}e_{ik}e_{jl}e_{kl})^{\frac{1}{2}} + (e_{ik}e_{il}e_{jk}e_{jl})^{\frac{1}{2}} = (e_{ij}e_{il}e_{jk}e_{kl})^{\frac{1}{2}}.$$

Since each index in (18) occurs twice in each term, when we substitute $e_{ij} = -\pi_{ij}/(2\pi_{ii}\pi_{jj}h_ih_j)$, from (17), the factors $\pi_{\mu\mu}$ and h_μ divide out. Thus we obtain the $\frac{1}{2}(n-3)(n-4) = \frac{1}{2}(g-2)(g-3)$ identities

$$(19) \qquad (\pi_{ij}\pi_{ik}\pi_{jl}\pi_{kl})^{\frac{1}{2}} + (\pi_{ik}\pi_{il}\pi_{jk}\pi_{jl})^{\frac{1}{2}} = (\pi_{ij}\pi_{il}\pi_{jk}\pi_{kl})^{\frac{1}{2}},$$

found originally by Poincaré [5]. These can also be wirtten

$$(20) \qquad (p_{ij}p_{ik}p_{jl}p_{kl})^{\frac{1}{2}} + (p_{ik}p_{il}p_{jk}p_{jl})^{\frac{1}{2}} = (p_{ij}p_{il}p_{jk}p_{kl})^{\frac{1}{2}} + O(t^5).$$

Thus we obtain in quite elementary fashion the result which Poincaré developed from the translational character of the manifold of zeros of the theta-functions. Of course, we are restricted in (20) to the case of our special symmetric Riemann surface S. Our derivation of (20) indicates, however, that our method is not limited to cases where all the elements of the matrix of the $p_{\mu\nu}$ tend to zero. Finally, note that using the approach of Section 2 we can show that the identities (19) are equivalent to the statement that the $(n-1)$-by-$(n-1)$ matrix $\| \gamma_{\mu\nu}\pi_{\mu\nu}^{-1} \|$ has rank 3. This deduction is based on the fact that the n-by-n matrix

$$\left\| \begin{matrix} b_ib_j, & 1 \\ 1, & 0 \end{matrix} \right\|$$

has rank 3 and the fact that the quantities h_i are arbitrary.

5. An alternate set of normal integrals. Let us return for a moment to the discussion of Section 1 and make a few remarks which have some bearing on the question of conformal type and the interpretation of the $p_{\mu\nu}$ as conformal moduli.

Let $w_{\mu\nu}(z)$ be such a conformal mapping of the multiply-connected region D upon the interior of a circular ring, cut along concentric circular slits, that C_μ and C_ν are carried into the inner and outer circles of the ring, respectively, while the remaining C_k correspond to the slits. Then one sees that, for example, the functions

$$W_\mu(z) = \tfrac{1}{2}\log \omega_{\mu n}(z), \qquad \mu = 1, \cdots, n-1,$$

yield a set of $g = n-1$ independent normal integrals of the first kind on the symmetric closed Riemann surface S generated by D. This set of integrals results when we choose the curves $C_1, \cdots, C_{n-1}$ to be the canonical cuts α_μ on S, rather than to be the cuts β_μ as in Section 1. The integrals $W_\mu(z)$ have a g-by-g period matrix $\| r_{\mu\nu} \|$ about the cuts β_ν, now taken to intersect the C_ν, and this set of quantities is given, evidently, in terms of the $p_{\mu\nu}$ by the relations

$$r_{\mu\nu} = (-1)^{\mu+\nu} \, | p_{ij} | / | p_{kl} | ,$$
$$k, l = 1, \cdots, g, \quad i, j = 1, \cdots, g, \quad i \neq \mu, j \neq \nu,$$

where the numerator is a $(g-1)$-by-$(g-1)$ determinant, the denominator a g-by-g determinant. Clearly $r_{\mu\nu} = r_{\nu\mu}$, and the $r_{\mu\nu}$ are, again, a special case of the invariants $a_{\mu\nu}$.

By the procedure outlined in this paper one can derive asymptotic identities among the periods $r_{\mu\nu}$ which are analogous to (10), (13) and (19). Indeed, the calculations for the $r_{\mu\nu}$ are even easier than those we have carried out for the $p_{\mu\nu}$. For example, in the case treated in Section 3, we have immediately

$$\operatorname{Re}\{W_\mu(z)\} = \tfrac{1}{2}(\log | z - a_\mu | - \log | 1 - \bar{a}_\mu z |) + \text{const.} + O(t),$$

so that

$$r_{\mu\nu} = \log | 1 - \bar{a}_\mu a_\nu | - \log | a_\mu - a_\nu | + O(t), \qquad \mu \neq \nu,$$
$$r_{\mu\mu} = \log(1 - | a_\mu |^2) - \log t\epsilon_\mu + O(t).$$

Thus our identities are in certain ways even more interesting when developed in terms of the matrix $\| r_{\mu\nu} \|$ rather than the matrix $\| p_{\mu\nu} \|$. It is for this reason that we call attention to these simple relationships here.

Note, then, that our asymptotic identities among the $r_{\mu\nu}$ in the case of

Section 3 are valid to within terms $O(t)$, and that the lowest order terms in the $r_{\mu\nu}$ $(\mu \neq \nu)$ do not tend to zero with t, so that the identities hold among the $r_{\mu\nu}$ themselves, except for the term $O(t)$. Thus, for example, the matrix

$$\left\| \begin{matrix} \gamma_{\mu\nu}/(e^{2r_{\mu\nu}}-1), & 1 \\ 1 \quad\quad , & -2 \end{matrix} \right\|$$

is of rank 4 for the domain of Section 3, if we neglect terms of order of magnitude $O(t)$. Hence we obtain to some extent further information about the identities among the $a_{\mu\nu}$ when we make the elementary change from the $p_{\mu\nu}$ to the $r_{\mu\nu}$.

For the limiting domain of Section 2, we have almost without calculation the formulas

$$W_\mu(z) = \tfrac{1}{2}\log(z - a_\mu) + \text{const.} + O(t), \qquad \mu = 1, \cdots, g,$$

whence

$$r_{\mu\nu} = -\log t\epsilon_n - \log | a_\mu - a_\nu | + O(t), \qquad \mu < \nu < n,$$

$$r_{\mu\mu} = -\log t^2\epsilon_n\epsilon_\mu + O(t), \qquad \mu < n.$$

Setting $q_{\mu\nu} = \lim\limits_{t\to 0} \{r_{\mu\nu} + \log t\}$, we find that our identities state that the matrix

$$\left\| \begin{matrix} \gamma_{\mu\nu}e^{-2q_{\mu\nu}}, & 1 \\ 1 \quad , & 0 \end{matrix} \right\|$$

has rank 4. This form is preferable to (8), (9) or (10) because it involves all asymptotic terms in $1/|\log t|$.

For the case developed in Section 4, the $r_{\mu\nu}$ and $p_{\mu\nu}$ behave alike, and nothing new is discovered.

So far we have made assumptions on the asymptotic behavior of D and have deduced asymptotic identities. Let us now reverse our procedure and demonstrate that our identities hold when we merely suppose that the quantities $r_{\mu\nu}$ have a prescribed limiting behavior. This will give our identities a more complete significance. We shall restrict ourselves to the situation described in Section 3, noting that the ideas apply in general.

Let us suppose, then, that we are given a domain D of connectivity n about which we know only that the invariants $r_{\mu\nu}$ have the form

$$(21) \qquad r_{\mu\nu} = -\delta_{\mu\nu}\log t + R_{\mu\nu} + O(t), \qquad \mu, \nu < n,$$

with $t \to 0$. We shall deduce that the matrix

$$\left\| \begin{matrix} \gamma_{\mu\nu}/e^{2R_{\mu\nu}} - 1), & 1 \\ 1 \quad , & -2 \end{matrix} \right\|$$

has rank 4, thus showing that our identities (13) of Section 3 hold as soon as it is known that the $p_{\mu\nu}$ tend to zero according to the rule

$$p_{\mu\nu} = \delta_{\mu\nu}/|\log t| + P_{\mu\nu}/|\log t|^2 + O(1/|\log t|^3), \qquad t \to 0,$$

which is equivalent to (21). We shall take the $R_{\mu\nu}$ to be distinct, except for the symmetry relation $R_{\mu\nu} = R_{\nu\mu}$.

We first notice that for a suitable constant γ_1 the function

$$e^{\gamma_1} w_{1n}(z) = \exp(2 W_1(z) + \gamma_1)$$

maps D upon the interior of the unit circle slit along circular arcs about the origin of radius $e^{-r_{\mu 1}}$, which correspond to the boundary components C_μ of D, $1 < \mu < n$, and with the circle of radius $e^{-r_{11}}$, which corresponds to C_1, deleted. We can assume, therefore, that D is, in fact, this image domain. We define s to be the maximum diameter of the components C_μ, $\mu < n$, of the boundary of our new model of the domain D.

It is easily seen that s tends to zero with t, since if this were not the case, (21) could not hold. Indeed, the annular region bounded by C_μ and C_n can be mapped conformally upon a circular ring with outer radius 1 and inner radius $R_\mu < e^{-r_{\mu\mu}}$ [3]. Hence $R_\mu = O(t)$, and therefore $s = O(t)$. It follows now readily from an application of Green's theorem to the functions

$$\operatorname{Re}\{W_\mu(z)\}, \qquad \log|z - \zeta| - \log|1 - \bar{\zeta} z|$$

that

$$\operatorname{Re}\{W_\mu(z)\} + \gamma_\mu/2 = \tfrac{1}{2}(\log|z - A_\mu| - \log|1 - \bar{A}_\mu z|) + O(t)$$

for z bounded away from C_μ, where A_μ is any point of C_μ. Thus if A_ν is any point of C_ν, we have

$$r_{\mu\nu} = \log|1 - \bar{A}_\mu A_\nu| - \log|A_\mu - A_\nu| + O(t).$$

As $t \to 0$, the points A_μ on the C_μ, $\mu < n$, tend to limiting positions a_μ, and we have

$$R_{\mu\nu} = \log|1 - \bar{a}_\nu a_\mu| - \log|a_\nu - a_\mu|, \qquad \mu < \nu < n.$$

Thus, indeed, our identities hold among the $R_{\mu\nu}$, as stated, and the matrix

$$\left\| \begin{matrix} \gamma_{\mu\nu}/(e^{2R_{\mu\nu}}-1), & 1 \\ 1 \quad , & -2 \end{matrix} \right\|$$

has rank 4. Similar procedures apply to our other cases, and we see that the identities of this paper do not depend upon assumptions as to the specific nature of the multiply-connected domain D.

It is also clear that the $r_{\mu\nu}$ determine the limit position of D uniquely up to reflection and linear fractional transformation.

STANFORD UNIVERSITY.

REFERENCES.

[1] L. Ahlfors and A. Beurling, "Invariants conformes et problèmes extrémaux II," *Comptes rendus du dixième congrès des mathématiciens Scandinaves*, Copenhague, 1947, pp. 341-351.

[2] P. R. Garabedian, "A remark on the moduli of Riemann surfaces of genus 2," *Bulletin of the American Mathematical Society*, to appear shortly.

[3] ——— and M. Schiffer, "Identities in the theory of conformal mapping," *Transactions of the American Mathematical Society*, vol. 65 (1949), pp. 187-238.

[4] A. Krazer, *Lehrbuch der Thetafunktionen*, Leipzig, 1903.

[5] H. Poincaré, "Remarques diverses sur les fonctions abéliennes," *J. de Math.*, sér. 5, vol. 1 (1895), pp. 219-314.

[6] F. Schottky, "Zur Theorie der Abelschen Functionen von vier Variabeln," *J. für Math.*, vol. 102 (1888), pp. 304-352.

ON THE POINT SPECTRA OF $y'' + (\lambda - q(x))y = 0$.*

By Göran Borg.

1. In several papers Wintner, Hartman and Wallach have studied the location of the continuous and discrete spectrum of

$$(L) \qquad y'' + (\lambda - q(x))y = 0,$$

$q(x)$ being continuous for $0 \leqq x < \infty$, and[1]

$$(R_\alpha) \qquad y(0)\cos\alpha + y'(0)\sin\alpha = 0, \qquad y(x) \subset L^2(0, \infty) \text{ for } \operatorname{Im}\lambda > 0.$$

The following is a result in the same direction.

If[2]

$$(1) \qquad \int_0^x |q(t)|\,dt \leqq k_0 \log x, \qquad x \to \infty,$$

then every point λ_n *in the point spectrum of* (L, R_α). *satisfies* $\lambda_n < k_0^2$.

There are equations (L), *satisfying* (1), *such that at least one eigenvalue* $\lambda = k^2$ *of* (L, R_α) *satisfies* $k > k_0 - \epsilon$ *(*ϵ *arbitrarily small).*

2. Proof of the first statement. Put $k^2 = \lambda\ (> 0)$. Then every solution $y(x, k)$ of (L) satisfies the standard inequality[3]

$$(2) \qquad k^2 y^2(x, k) + y'^2(x, k) \geqq \text{const.} \exp\left(-k^{-1}\int_0^x |q(t)|\,dt\right)$$

for some const. > 0. According to (1), the exponential multiplying the const. is minorized by $x^{-k_0/k}$. Hence, if x_n denotes a point at which $y'(x, k) = 0$, then, according to (2),

$$y^2(x_n, k) \geqq \text{const. } x_n^{-k_0/k}.$$

---

* Received May 29, 1950.

[1] If (R_α) does not alone define an eigenvalue problem, one more condition at $x = 0$ must be given. This is not the case here, for condition (1) implies limit-point behavior (no L^2-solution for $k > k_0$); cf. also the general results in this direction of Hartman and Wintner, *American Journal of Mathematics*, vol. 71 (1949), p. 206.

[2] There are also connections with the paper of Hartman and Wintner in *American Journal of Mathematics*, vol. 70 (1948), pp. 295-308 concerning the non-degeneracy of the wave equation.

[3] Actually, (2) is a particular case of an inequality used by Liapounoff in case of systems; for references, cf. Picard's Traité d'Analyse, 3rd ed. (1928), vol. 3, p. 385.

Suppose now that

$$\liminf nx_n^{-1} = r > 0 \qquad (n \to \infty) \tag{3}$$

and

$$\int_{x_n-\delta}^{x_n+\delta} y^2(x, k)\,dx > Cy^2(x_n, k) \tag{4}$$

(δ fixed, C fixed for any n). Then if $k \geqq k_0$,

$$\int_0^\infty y^2(x, k)\,dx = \infty. \tag{5}$$

From this the statement would follow.

In fact, (3) holds true and (4) is satisfied for a subset of indices $n = m_p$ $(p = 1, 2, 3, \cdots)$, dense enough to imply (5). We begin with the proof of (3), and show that the relation holds if x_n represents the n-th zero of $y(x, k)$. In proving $r > 0$, we can suppose $q(x) \geqq 0$; for, replacing $q(x)$ in (L) by $|q(x)|$, we get a set of zeros x'_n and a lim inf $= r'$ satisfying $x'_n \geqq x_n$, that is, $r' \leqq r$. In order to avoid confusion, let ξ_p denote the p-th zero of $y(x, k)$ for $x > 0$. Since now $q(x) \geqq 0$, we have $\xi_p - \xi_{p-1} = l_p \geqq \pi k^{-1}$. Then we can apply the Hilfssatz 2 in an earlier paper of the author which here takes the following form:[4]

If $l_p = \xi_p - \xi_{p-1} \geqq \pi k^{-1}$, then

$$\int_{\xi_{p-1}}^{\xi_p} |\, y''y^{-1} + k^2 \,|\, dx \geqq k^2 l_p (1 - \pi l_p^{-1} k^{-1}).$$

From this we get $\int_{\xi_1}^{\xi_p} |\,q(t)\,|\,dt \geqq k^2(\xi_p - \xi_1 - (p-1)\pi k^{-1})$. Hence from (1), $p\xi_p^{-1} \geqq k\pi^{-1} + o(1)$, that is, $\liminf p\xi_p^{-1} \geqq k\pi^{-1}$. Hence the same inequality will hold for any function $q(x)$ satisfying (1). But $x_2 < \xi_2$, $x_3 < \xi_3, \cdots, x_p < \xi_p, \cdots$, hence $px_p^{-1} > p\xi_p^{-1}$ (after suitable numbering), which implies (3).

The proof of (5) is now simple. In fact, (3) and (1) imply that

$$\int_{x_n-\delta}^{x_n+\delta} |\,q(t)\,|\,dt > \epsilon$$

(δ fixed) will hold only for certain indices $n = m_q$ $(q = 1, 2, 3, \cdots)$ such that $m_q < n$ implies $q = O(\log n)$. Let the indices which correspond to the reversed inequality be denoted m_p $(p = 1, 2, 3, \cdots)$. Since, for $m = m_p$,

---

[4] G. Borg, Arkiv för Mat. Astr. och Fys., vol. 31, no. 1 (1944), p. 10.

$$y(x, k) = y(x_m, k) \cos k(x - x_m) + \int_{x_m}^{x} \sin k(x - t) q(t) y(t, k) k^{-1} dt,$$

we obtain

$$y(x, k) = y(x_m, k) \cos k(x - x_m)(1 + \epsilon(x)), \qquad |\epsilon(x)| < \tfrac{1}{2}$$

in a δ-neighborhood of x_m ($m = m_p$, δ independent of x, x large). Thus (4) follows for $n = m_p$ $(p = 1, 2, 3, \cdots)$ and hence, if $k \geqq k_0$,

$$\int_0^n y^2(x, k)\, dx \geqq C \sum_{m_p \leqq n} m_p{}^{-k_0/k} \geqq C \sum_{O(\log n)}^{n} p^{-k_0/k} \to \infty, \qquad n \to \infty.$$

3. In order to prove the second part of the theorem, we have to construct a function $y(x, k)$ satisfying

$$\int_0^\infty y^2(x, k)\, dx < \infty \tag{6}$$

and an equation (L) with $\lambda = k^2 > (k_0 - \epsilon)^2$ and with (1). This will be done with the aid of sine functions with corners, as follows. Define $y(x, k) = A_n \sin(kx - \eta_n)$ for $t_{n-1} \leqq x \leqq t_n$ and so on, so that

$$\begin{aligned} y(t_n - 0, k) &= y(t_n + 0, k) = y(t_n, k), \\ y'(t_n + 0, k) - y'(t_n - 0, k) &= y(t_n, k)\delta_n. \end{aligned} \tag{7}$$

Then $y(x, k)$ will satisfy the generalized equation (L):

$$dy' + k^2 y - y\, dQ(x) = 0, \tag{L'}$$

where $Q(x)$ is a step-function, constant except for $x = t_n$, where $\int_{t_n} dQ(x) = \delta_n$. Let there be only one corner between two successive zeros of $y(x, k)$, denote the zeros, as above, with ξ_n and let $\xi_n < t_n < \xi_{n+1}$. If $\alpha_n = k(t_n - \xi_n)$, the relations (7) give

$$\begin{aligned} A_{n+1}{}^2 &= A_n{}^2(1 + \delta_n k^{-1} \sin 2\alpha_n + \delta_n{}^2 k^{-2} \sin^2 \alpha_n) \\ \cot(kt_n - \eta_{n+1}) &= \cot(kt_n - \eta_n) + \delta_n k^{-1}. \end{aligned} \tag{8}$$

4. It is now a simple matter to construct an equation (L') satisfying the relation

$$\int_0^x |dQ(x)| \leqq k_0 \log x$$

(corresponding to (1)), which for a $\lambda^{\frac{1}{2}} > k_0 - \epsilon$ has a solution satisfying (6). It only remains to round off the corners of the solution $y(x, k)$ in order to get an equation (L) with properties required in the theorem. Put $(k_0 - \epsilon')\log x = g(x)$ $(\epsilon' < \epsilon)$ and $\delta_1 = g(t_1)$, $\delta_2 = g(t_2) - g(t_1), \cdots$. $\delta_n = g(t_n) - g(t_{n-1})$, the t_n being the corners of $y(x, k)$ which we determine in the following way:

Choose $y = A_1 \sin(kx - \eta_1)$ such that (R_a) is satisfied for $x = 0$. If ξ_1 is the first zero of $y(x, k)$ which is > 0, choose a t_1 such that $\alpha_1 = k(t_1 - \xi_1) = 3/4\pi$. Now let $y(x, k)$ be defined by the expression above up to $x = t_1$. Then apply relations (8) with the numbers δ_1 and α_1 defined above, which will give us A_2 and η_2, determining the function $y(x, k)$ for $x > t_1$. If $y(x, k)$ is defined up to the n-th zero ξ_n, choose t_n such that $\alpha_n = k(t_n - \xi_n) = 3/4\pi$ and put $y(x, k) = A_n \sin(kx - \eta_n)$ for $t_{n-1} \leqq x \leqq \xi_n$. Let the same definition hold for $t_{n-1} \leqq x \leqq t_n$. Apply (8) with the δ_n and α_n defined above. This determines $y(x, k) = A_{n+1} \sin(kx - \eta_{n+1})$ for $x > t_n$. Let ξ_{n+1} be the first zero of this function which is $> t_n$ and repeat the preceding procedure.

This construction implies that $t_{n+1} - t_n = O(1)$ $(n \to \infty)$, hence

$$\int_0^\infty y^2(x, k)\,dx \leqq C \sum A_n^2 (t_n - t_{n-1}) \leqq \text{const.} \sum A_n^2.$$

But according to (8) and the choice of δ_n and α_n we have, since $\delta_p = (k_0 - \epsilon') \log(t_p/t_{p-1}) \to 0$ when $p \to \infty$,

$$A_{n+1}^2 = \prod_1^n (1 - \delta_p k^{-1} + \delta_p^2 (2k^2)^{-1}) \leqq \text{const.} \exp(-k^{-1}(g(t_n) - g(t_1)))$$

$\leqq \text{const.}\ t_n^{-(k_0 - \epsilon')/k}$. Since $t_n - t_{n-1} = O(1)$, we have $t_n = O(n)$, which implies (6) for any $k < k_0 - \epsilon'$. Furthermore

$$(9) \qquad \int_0^x |dQ(x)| = \sum_{t_n \leqq x} \delta_n = (k_0 - \epsilon')\log t_n \leqq k_0 \log x.$$

Since $\epsilon' < \epsilon$, this is the proposition, with an equation (L') instead of an equation (L).

It remains to replace the $y(x, k)$, defined above, by a twice continuously differentiable function in such a way that (6) still holds (for a $k = \lambda^{\frac{1}{2}} > k_0 - \epsilon$) and the inequality of (9) turns into (1). This can be done as, for instance, in the paper of the author in vol. 71 (1949), p. 69 of this JOURNAL. The substitution

$$u = y(x)\,(\sin(kx - \eta_n))^{-1}, \qquad v = -\cot(kx - \eta_n)$$

transforms the corner between the two sine functions at $x = t_n$ into one between straight lines, the case in the paper just quoted. Only simple computations are necessary to establish that the "rounding-off," made there at one corner, say t_n, can be made uniformly for all n (this being the only new point) and in such a way that the above requirements become satisfied.

5. This theorem, giving a certain "best possible" result of the penetrating of the point spectrum into the continuous spectrum of (L, R_a), has also another aspect. It is a theorem on the growth of the solutions of (L) under condition (1), and it is interesting to notice that equations of the kind (L') have a certain "extremal" property in this connection; cf. also the papers by the author cited above.

The same remark applies to the theorem of Levinson[5] on the growth of the solutions. He obtains a constant $1/\pi$, characteristic of the growth, which in fact can be replaced by the precise constant $\frac{1}{2}$ by means of a construction similar to the one used above in Sections 3-4.

THE UNIVERSITY OF UPPSALA, SWEDEN.

---

[5] N. Levinson, *Duke Mathematical Journal*, vol. 8 (1941), pp. 1-10.

ON COMMUTATORS OF BOUNDED MATRICES.*

By C. R. Putnam.

1. It was shown by Wintner [5] that if A and B are bounded Hermitian matrices then the commutator C of A and B, that is

$$C = AB - BA, \tag{1}$$

cannot be a constant multiple, other than zero, of the unit matrix; cf. also Rellich [2]. It is clear, however, from the argument given in [5], that this conclusion remains valid, provided only that A and B are bounded, whether or not A and B are Hermitian; see Appendix (i) below. This extension of the theorem was formulated and proved by Wielandt [3], using a method independent of that in [5], for the case that A and B are bounded operators in a linear space (not necessarily a Hilbert space) with a linear metric.

The question was raised by Professor Wintner (oral communication) whether the mere assumptions that A and B are bounded might imply the stronger result that the greatest lower bound of the collection of numbers $|x^* Cx|$, where the vector x ranges over the Hilbert unit sphere, is zero; that is, (1) satisfies

$$\operatorname*{g.l.b.}_{\|x\|=1} |x^* Cx| = 0. \tag{2}$$

(The norm $(x^* x)^{\frac{1}{2}}$ of x will be denoted by $\| x \|$; for terminology, cf. [1].) That (2) does hold in case A and B are finite matrices is easily verified by the following argument. Let A be a finite matrix and denote its trace by $\mathrm{tr}(A)$. Since, for any two finite matrices, A and B, of the same order, $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$ and $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, it follows from (1) that $\mathrm{tr}(C) = 0$. According to Hausdorff (cf. [4], p. 34) the collection of numbers $W: x^* Cx$, where $\| x \| = 1$, is a convex set. Since each of the diagonal elements of C is in the set W (the n-th diagonal element of C being equal to $x^* Cx$ if x is the vector whose n-th component is 1 and whose remaining components are 0), it follows that $\mathrm{tr}(C)/n$ is in W and consequently (2) holds.

Whether or not (2) is true for arbitrary bounded A and B will remain an open question. The following theorem will, however, be proved:

---

* Received June 14, 1950.

(*) *Let A and B denote bounded matrices of the Hilbert space consisting of vectors x. For every real number $\lambda \neq 0$, let $M(\lambda)$ denote the greatest point of the spectrum of the Hermitian matrix*

$$D(\lambda) = \lambda(AA^* - A^* A) - \lambda^{-1}(BB^* - B^* B). \tag{3}$$

If $\eta \geqq 0$ *and*

$$\eta \geqq \underset{\lambda \neq 0}{\text{g. l. b.}}\, M(\lambda), \tag{4}$$

then (1) *satisfies*

$$\underset{\|x\|=1}{\text{g. l. b.}}\, | x^* Cx | \leqq \tfrac{1}{2}\eta. \tag{5}$$

The following definitions will be used in what follows. A bounded matrix B will be called non-singular if B possesses a *unique bounded* right inverse B^{-1}; in which case, B^{-1} is also the unique bounded left inverse of A. A bounded Hermitian matrix H will be called definite [semi-definite] if either l. u. b. $x^* Hx < 0$ [$\leqq 0$] or g. l. b. $x^* Hx > 0$ [$\geqq 0$], where $\| x \| = 1$.

As a consequence of (*) it is possible to derive the following

Corollary. *If either of the matrices $AA^* - A^* A$ or $BB^* - B^* B$ is semi-definite, then* (1) *satisfies* (2).

It is clear from (3) and from the fact that λ can be any non-zero value that is is sufficient to prove the Corollary in the case that $AA^* - A^* A$ is negative semi-definite. Let the norm $| A |$ of a bounded matrix A be defined by

$$| A | = \text{l. u. b.}\, | y^* Ax |, \qquad \| x \| = \| y \| = 1.$$

Then (3) shows that, for $\lambda > 0$ and $\| x \| = 1$,

$$x^* D(\lambda)x \leqq -\lambda^{-1}x^*(BB^* - B^* B)x,$$

and therefore, since $| BB^* - B^* B | \leqq 2 | B |^2$,

$$x^* D(\lambda)x \leqq 2 | B |^2\lambda^{-1}.$$

The matrix $D(\lambda)$ is Hermitian; hence $M(\lambda) = \text{l. u. b.}\, x^* D(\lambda)x$, where $\| x \| = 1$, so that $M(\lambda) \leqq 2 | B |^2\lambda^{-1}$. Since λ is an arbitrary positive number, (4) holds for $\eta = 0$ and (2) follows from (*).

Remark. Clearly, if A is normal, then $AA^* - A^* A$ (the zero matrix) is semi-definite. It is noteworthy that the converse of this statement is not true in general, but does hold if A is a finite matrix. That is, *there do not exist finite, but there do exist bounded, matrices for which $AA^* - A^* A$ is semi-definite and A is not normal*; cf. Appendix (ii).

2. *Proof of* (*). For any pair of real numbers $\lambda \neq 0$ and θ define the bounded matrix F by

$$F = |\lambda|^{\frac{1}{2}} e^{i\theta} A + |\lambda|^{-\frac{1}{2}} B^*, \text{ so that } F^* = |\lambda|^{\frac{1}{2}} e^{-i\theta} A^* + |\lambda|^{-\frac{1}{2}} B.$$

It is easily verified that

$$(6) \qquad FF^* - F^* F = |\lambda| (AA^* - A^* A) - |\lambda|^{-1} (BB^* - B^* B) + H,$$

where $H = e^{i\theta} C + e^{-i\theta} C^*$. If $D(\lambda)$ is defined by (3), relation (6) can be expressed as

$$(7) \qquad (\operatorname{sgn} \lambda)(FF^* - F^* F) = D(\lambda) + (\operatorname{sgn} \lambda) H,$$

and consequently,

$$(8) \qquad (\operatorname{sgn} \lambda) x^* (FF^* - F^* F) x = x^* D(\lambda) x + 2 (\operatorname{sgn} \lambda) \Re (e^{i\theta} x^* Cx).$$

Suppose (*) is false; that is, suppose that there exists a real number $\eta \geqq 0$ such that (4) holds and (5) fails to hold. Thus $\tau > \frac{1}{2}\eta$, where

$$(9) \qquad \tau = \operatorname*{g.l.b.}_{\|x\|=1} | x^* Cx | .$$

It is known (cf. [4], p. 34) that the closure W of the collection of numbers $x^* Cx$, where $\| x \| = 1$, is a convex set. (If W_n denotes the set of numbers $x^* C_n x$, where $\| x \| = 1$ and C_n is the finite n-th section matrix of C, then W_n is contained in W_{n+1} and W is the closure of the set ΣW_n.) Since $\tau > 0$, it is possible to choose real values θ so that either $\Re(e^{i\theta} x^* Cx) \geqq \tau$ or $\Re(e^{i\theta} x^* Cx) \leqq -\tau$ for all vectors x satisfying $\| x \| = 1$. (The set $We^{i\theta}$ is merely the set W rotated through an angle θ.) In virtue of (4), it is possible to find a real number $\lambda \neq 0$ such that $M(\lambda) < 2\tau$; for this value λ, choose a real θ so that $(\operatorname{sgn} \lambda) \Re(e^{i\theta} x^* Cx) \leqq -\tau$. For the above pair λ and θ it is seen that, since $D(\lambda)$ is Hermitian, the right side of (8) is not greater than $-2\tau + M(\lambda)$ for all vectors x satisfying $\| x \| = 1$. That is, the operator $D(\lambda) + (\operatorname{sgn} \lambda) H$ is negative definite. It will now be shown that if F is bounded, then $FF^* - F^* F$ cannot be definite.

It can be supposed that F is non-singular (otherwise, replace F by G, where G differs from F by an appropriate constant multiple of the unit matrix, and note that $GG^* - G^* G = FF^* - F^* F$). Since $FF^* = F(F^* F) F^{-1}$, the matrices FF^* and $F^* F$ have the same spectrum; cf. [5].

It is seen that $x^* (FF^* - F^* F) x = x^* FF^* x - x^* F^* Fx$. Let ϵ denote an arbitrary positive number, and λ_0 the least point of the spectrum of FF^* and $F^* F$. Clearly, both FF^* and $F^* F$ are Hermitian. Thus there exists a vector y, $\| y \| = 1$, such that $y^* FF^* y < \lambda_0 + \epsilon$; moreover, $x^* F^* Fx \geqq \lambda_0$

for all x with $\| x \| = 1$, and consequently $y^*(FF^* - F^*F)y < \epsilon$. Since the positive number ϵ is arbitrary, $FF^* - F^*F$ cannot be positive definite. Similar reasoning when λ_0 is the greatest point of the spectrum of FF^* and F^*F shows that $FF^* - F^*F$ cannot be negative definite. These results are in contradiction with (7) and the previous result that $D(\lambda) + (\operatorname{sgn} \lambda)H$ is negative definite for some real $\lambda \neq 0$. Accordingly, a contradiction to the assumption that (*) is false is obtained and the proof of (*) is complete.

Appendix.

(i) Since (1) remains unchanged if A is replaced by a matrix differing from A by a constant multiple of the unit matrix, it can be supposed that A is non-singular. Since $AB = A(BA)A^{-1}$, it follows that AB and BA have identical spectra. Suppose, if possible, that (1) is a constant multiple $\sigma \neq 0$ of the unit matrix; then the spectrum of AB is the set obtained by translating the spectrum of BA by σ. Hence the spectrum of AB (or BA) is unbounded; since, however, AB and BA are bounded, their spectra are bounded. This contradiction shows that C cannot be a non-zero constant multiple of the unit matrix; cf. [5], where the above argument is used.

(ii) Let A denote a finite matrix for which the Hermitian matrix $AA^* - A^*A$ is semi-definite. Without loss of generality it can be supposed that the latter matrix is non-negative definite and that consequently its eigenvalues are non-negative. Since $\operatorname{tr}(AA^* - A^*A) = 0$, it follows that all these eigenvalues are zero. Hence $AA^* - A^*A$ is equivalent (in fact, unitarily equivalent) to the zero matrix and consequently is itself the zero matrix; that is, A is normal. On the other hand, there exist bounded matrices A such that $AA^* - A^*A$ is semi-definite and A is not normal. Such a matrix $A = (a_{ij})$ is given by $a_{ii+1} = 1$, $a_{ij} = 0$, $j \neq i + 1$. It is easily verified that the elements of $AA^* - A^*A = (\alpha_{ij})$ are all zero except α_{11} $(= 1)$. Actually the matrix A is singular (since A^* is a right but not a left hand reciprocal) but a non-singular matrix B for which $BB^* - B^*B = AA^* - A^*A$, is easily constructed; cf. (i) above.

(iii) It follows from (*) that if $\tau > 0$, where τ is defined by (9), then $\tau \leqq \frac{1}{2} \operatorname*{l.\,u.\,b.}_{\|x\|=1} x^* D(\lambda)x$, for every fixed real $\lambda \neq 0$. Define a and b by

$$a = \operatorname*{l.\,u.\,b.}_{\|x\|=1} | x^*(AA^* - A^*A)x | \quad \text{and} \quad b = \operatorname*{l.\,u.\,b.}_{\|x\|=1} | x^*(BB^* - B^*B)x | .$$

Clearly, if $\tau \geqq 0$, then $\tau \leqq \frac{1}{2}(\lambda a + b\lambda^{-1})$ holds for all $\lambda > 0$. Suppose that $a > 0$ and $b > 0$. Then $\frac{1}{2}(\lambda a + b\lambda^{-1})$ has for $0 < \lambda < \infty$ the minimum value $(ab)^{\frac{1}{2}}$ so that

$$\tau \leqq (ab)^{\frac{1}{2}}. \tag{10}$$

If $a = 0$, it follows from the definition of a that A is normal; if $b = 0$, then B is normal. Thus, if either $a = 0$ or $b = 0$, the Corollary of (*) shows that $\tau = 0$ so that (10) is still valid. These results may be summarized in the inequality

$$(11) \quad (\underset{\|x\|=1}{\text{g. l. b.}} \, | x^*(AB - BA)x |)^2 \leq (\underset{\|x\|=1}{\text{l. u. b.}} \, | x^*(AA^* - A^*A)x |)(\underset{\|x\|=1}{\text{l. u. b.}} \, | x^*(BB^* - B^*B)x |),$$

which is valid for any pair of bounded matrices A and B.

It remains undecided whether (11) is trivial in the sense that the left side is zero for every such pair of matrices.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] B. v. Sz. Nagy, *Spektraldarstellung linearer Transformationen des Hilbertschen Raumes* (*Ergebnisse der Mathematik*), Berlin, 1942.

[2] F. Rellich, "Der Eindeutigkeitssatz für die Lösungen quantenmechanische Vertauschungsrelationen," *Göttinger Nachrichten*, 1946, pp. 107-115.

[3] H. Wielandt, "Ueber die Unbeschränktheit der Operatoren der Quantenmechanik," *Mathematische Annalen*, vol. 121 (1949), p. 21.

[4] A. Wintner, *Spektraltheorie der unendlichen Matrizen*, Leipzig, 1929.

[5] ———, "The unboundedness of quantum-mechanical matrices," *Physical Review*, vol. 71 (1947), pp. 738-739.

ON THE PROBLEMS OF GEODESICS IN THE SMALL.*

By Philip Hartman and Aurel Wintner.

1. In the positive definite line element

$$ds^2 = g_{ik} du_i du_k, \qquad (u_1, u_2) = (u, v), \tag{1}$$

let the coefficient functions $g_{ik} = g_{ik}(u, v)$ be defined and of class C^1 in a vicinity of a point, say $(u, v) = (0, 0)$. Then the Christoffel symbols $\Gamma^i{}_{jk}$ exist and are continuous functions of (u, v). Hence the differential equations

$$u_i'' + \Gamma^i{}_{jk} u_j' u_k' = 0 \tag{2}$$

are meaningful and define the geodesics belonging to (1). The term "geodesic" below will always refer to a solution of (2) (rather than to an arc minimizing the arc-length integral belonging to (1)).

The question of whether or not an initial point and direction determine a unique geodesic (locally) has been answered in the negative. ([5], § 2). Despite this fact, the following positive assertion will turn out to be true:

(I) *Let* (1) *be positive definite, with coefficients* $g_{ik}(u, v)$ *of class* C^1, *in a vicinity of* $(u, v) = (0, 0)$. *Then, to every sufficiently small* $\epsilon > 0$, *there belongs a* $\delta > 0$ *having the property that every point* $(u, v) \neq (0, 0)$ *of the circle* $C_\delta\colon u^2 + v^2 \leqq \delta^2$ *can be joined with* $(0, 0)$ *by a geodesic contained in the circle* C_ϵ.

2. The following questions immediately arise:

(i) Under the assumptions of (I), is the geodesic supplied by (I) unique?

(ii) If $(u, v) \neq (0, 0)$ is a point of C_ϵ, then, according to Hilbert ([6], p. 186; cf. [3], pp. 422-427), there exists in C_ϵ a rectifiable arc joining (u, v) with $(0, 0)$, for which the arc-length (in the metric (1)) is the least among all such arcs. Is a minimizing arc unique if (u, v) is sufficiently near to $(0, 0)$?

(iii) Corresponding to a given direction through $(0, 0)$, are there points (u, v), arbitrarily near to $(0, 0)$, such that a minimizing arc, contained in C_ϵ and joining $(0, 0)$ with (u, v), has the assigned initial direction?

---

* Received January 5, 1950.

(iv) Are the answers to (i), (ii) and/or (iii) affirmative when every direction at $(0,0)$ determines a unique geodesic (locally) through $(0,0)$?

(v) Conversely, if the answers to (i), (ii) and/or (iii) are affirmative, does the point $(0,0)$ and a direction through $(0,0)$ determine a unique geodesic (locally)?

(vi) Must a geodesic furnish the minimal distance (in the sense of Hilbert, cf. (ii)) between some pair of points on it?

(II) *Under the assumptions of* (I), *the answer to each of the questions* (i)-(vi) *is in the negative.*

Remark. It will be seen that the answers to (i)-(vi) remain negative even if the Christoffel symbols Γ^i_{jk}, instead of being just continuous as in (I), (II), are required to satisfy a uniform Hölder condition of any given order $\alpha < 1$. But if $\alpha = 1$, the situation changes entirely:

(II bis) *If, in addition to the assumptions of* (I), *the partial derivatives* $\partial g_{ik}/\partial u_j$ *are assumed to satisfy a uniform Lipschitz condition, then the answer to each of the questions* (i)-(vi) *is in the affirmative.*

3. More than (II bis) is contained in the following theorem:

(III) *Under the assumptions of* (II bis), *there exists, on a sufficiently small circle* $\mathfrak{C}: (0 \leqq r \leqq a,\ 0 \leqq \phi < 2\pi)$, *a pair of functions*

$$(3) \qquad u = u(r, \phi), \qquad v = v(r, \phi)$$

having the following properties: Both functions (3) *are of class* C^1 *on* $\mathfrak{C}$, *vanish at the center of* $\mathfrak{C}$, *and* (3) *is, for every fixed* ϕ, *a geodesic, on which* r *is the arc-length. Furthermore,* (3) *represents a topological mapping of* $\mathfrak{C}$ *on a neighborhood of* $(u, v) = (0, 0)$ *and the Jacobian* $\partial(u, v)/\partial(r, \phi)$ *vanishes only at the center of* $\mathfrak{C}$. *Finally,* (3) *transforms* (1) *into the geodesic normal form*

$$(4) \qquad ds^2 = dr^2 + g d\phi^2,$$

where $g = g(r, \phi)$ *is continuous on* $\mathfrak{C}$ *and is positive except at the center of* $\mathfrak{C}$.

The classical methods prove (III) under the assumption that the functions g_{ik} are of class C^2, or even of class C^3, instead of being, as in (III), only of class C^1 with partial derivatives $\partial g_{ik}/\partial u_j$ satisfying a uniform Lipschitz condition. Correspondingly, the first point in (III) is that the fundamental field construction, classically based on the class C^2 (cf. [4], pp. 408-409), can be carried out under the less severe assumptions of (II bis).

The situation goes even further in the existence proof for the Gauss normal form (4). In fact, the usual proof has to assume, to this end, that the functions g_{ik} are of class C^3 (cf., for instance, [2], p. 101, where the deduction of (4) is based on the calculation of the Christoffel symbols in terms of the new coordinates r, ϕ; a calculation involving a differentiation, the legitimacy of which is anything but obvious under the C^2-assumption).

Clearly, the deduction of (4) under the relaxed assumptions of (III) is equivalent to the deduction, under such assumptions, of Gauss's theorem on the transversals of geodesics. Correspondingly, the point will be that Gauss's theorem can be deduced under assumptions which are less severe than in the usual proofs.

4. Proof of (I). Let $\epsilon > 0$ be so small that the g_{ik} are of class C^1 in the closed circle $C_\epsilon : u^2 + v^2 \leqq \epsilon^2$. Let M, m be a pair of constants satisfying

$$M \geqq g_{ik}(u, v)\xi_i\xi_k/(\xi_1{}^2 + \xi_2{}^2) \geqq m > 0 \tag{5}$$

for every (u, v) in C_ϵ. Choose δ so small that $m\epsilon > M\delta$. If (u, v) is a point of C_δ, let $\Gamma = \Gamma(u, v)$ be a rectifiable arc in C_ϵ joining $(0, 0)$ with (u, v) and minimizing the arc-length, in the metric (1), among all such arcs. Then no point of Γ is on the boundary of C_ϵ. For the length of the arc joining $(0, 0)$ with a point of the boundary of C_ϵ is at least $m\epsilon$, while the length of the line segment joining $(0, 0)$ with (u, v) is at most $M\delta < m\epsilon$.

In order to prove (I), it will be shown that Γ is a geodesic (which, in view of (2), implies that Γ is an arc of class C^2). Let L be the length of Γ and let $u = u(s)$, $v = v(s)$, where $0 \leqq s \leqq L$, be the parametric representation of Γ with s as the arc-length (in the metric (1)). Then $u(s)$, $v(s)$ are absolutely continuous and

$$g_{ik}(u(s), v(s))u_i'(s)u_k'(s) = 1, \qquad (u_1, u_2) = (u, v), \tag{6}$$

for all s on $0 \leqq s \leqq L$, with the possible exception of an s-set of measure 0. Du Bois-Reymond's derivation of the Euler-Lagrange equations shows that, in virtue of (6), there exist constants c_1, c_2 for which

$$g_{jk}u_k' - \tfrac{1}{2}\int_0^s (\partial g_{ik}/\partial u_j)u_i'u_k'dt = c_j \qquad (j = 1, 2)$$

holds almost everywhere. Hence, if (g^{ik}) denotes the matrix inverse to (g_{ik}), then

$$u_m' = g^{jm}\left(\tfrac{1}{2}\int_0^s (\partial g_{ik}/\partial u_j)\, u_i' u_k'\, dt + c_j\right) \qquad (m = 1, 2)$$

holds almost everywhere. If $u_1'(s)$, $u_2'(s)$ are defined for *all* s on $0 \leqq s \leqq L$ by the last formula line, then $u_1'(s)$, $u_2'(s)$ are continuous, while the functions $u_1(s) = \int_0^s u_1'(t)\,dt$, $u_2(s) = \int_0^s u_2'(t)\,dt$ coincide with the functions $u(s)$, $v(s)$ occurring in the above parametric representation of Γ.

Hence Γ is of class C^1. Standard theorems in the calculus of variations imply that Γ is of class C^2 and, therefore, a geodesic (in the sense of (2)). This completes the proof of (I).

5. The first counter-example. In the strip $|u| < 1$ of the (u, v)-plane, put

(7) $$g_{11} = 1, \quad g_{12} = 0, \quad g_{22} = 1 - |u|^\lambda, \text{ where } 1 < \lambda < 2.$$

These functions g_{ik} are of class C^1. The corresponding equations (2) become

(8) $$u'' + \tfrac{1}{2}\lambda(\operatorname{sgn} u)\,|u|^{\lambda-1}v'^2 = 0 \text{ and } ((1 - |u|^\lambda)v')' = 0.$$

The last equation gives

(9) $$(1 - |u|^\lambda)v' = c = \text{const.}$$

If $c = 0$, then $v = \text{Const.}$, while u is a linear function of t; hence the parameter curves $v = \text{Const.}$ are geodesics. If $c \neq 0$, then $v' \neq 0$ along the geodesic, so that v is strictly monotone along all geodesics except when $v = \text{Const.}$

If the independent variable is chosen to be the arc-length, then (6) is applicable. In view of (7), this means that $u'^2 + c^2(1 - |u|^\lambda)^{-1} = 1$, that is,

(10) $$u' = \pm(1 - c^2(1 - |u|^\lambda)^{-1})^{\frac{1}{2}}.$$

The case $c^2 = 1$ leads to the geodesic $u \equiv 0$.

Since every choice of + or — and every choice of c on the interval $-1 \leqq c \leqq 1$ in (10) lead to a (locally) unique solution of (10), it now follows from (10) and (9) that *every direction through* $(u, v) = (0, 0)$ *determines a unique geodesic.*

6. It will now be shown that the example (7) proves all but the fifth of the six negations in (II).

Ad (i). Consider a geodesic with an initial direction near $(u', v') = (0, 1)$; this corresponds to a value of c in (10) which is less than 1 but is close to 1. For convenience, it will be assumed that the $+$ sign is chosen in (10) for $u = 0$. Let $c = (1 - \epsilon)^{\frac{1}{2}} > 0$, where $\epsilon > 0$, and let the corresponding geodesic be called $\Gamma_{+\epsilon}$. Then $u' > 0$ at $u = 0$, and u' decreases until it becomes 0 at $u = \epsilon^{1/\lambda}$. Subsequently, u decreases, since otherwise the right-hand side of (10) would become imaginary (so that the $-$ sign must then be chosen in (10)). If $v = v_0(\epsilon)$ denotes the value of v at the point $u = \epsilon^{1/\lambda}$ of $\Gamma_{+\epsilon}$, then it is clear from the fact that v does not occur explicitly in (7), that $\Gamma_{+\epsilon}$ is symmetric with respect to the line $v = v_0$. Hence the geodesic $\Gamma_{+\epsilon}$ joins the point $(0, 0)$ with $(0, 2v_0)$. Clearly, these two points are joined by the geodesic $u \equiv 0$ also.

It will now be shown that $v_0(\epsilon) \rightarrow 0$ as $\epsilon \rightarrow 0$. This will imply that, in any neighborhood C_ϵ of $(u, v) = (0, 0)$, there are points $(0, 2v_0)$ which can be joined with $(0, 0)$ by more than one geodesic contained in C_ϵ. Hence, (7) shows that the answer to (i) is in the negative.

Let s denote the arc-length (in the metric (7)) on $\Gamma_{+\epsilon}$ measured from $(u, v) = (0, 0)$. If $\epsilon < \frac{1}{2}$, it follows from (9), with $c = (1 - \epsilon)^{\frac{1}{2}}$, that $0 < v' \leqq 2$; so that $v(s) \leqq 2s$ if $s > 0$. In particular, if $s = s_0 > 0$ is the (first) value of s satisfying $u(s) = \epsilon^{1/\lambda}$, then $v_0 \leqq 2s_0$. Hence, in order to prove that $v_0(\epsilon) \rightarrow 0$, as $\epsilon \rightarrow 0$, it is sufficient to show that $s_0 = s_0(\epsilon) \rightarrow 0$.

Clearly, (10) implies that

$$s_0 = \int_0^a \{1 - (1 - \epsilon)(1 - u^\lambda)^{-1}\}^{-\frac{1}{2}} du, \text{ where } a = \epsilon^{1/\lambda}.$$

A simplification of the integrand shows that $s_0 = \int_0^a (1 - u^\lambda)^{\frac{1}{2}} (\epsilon - u^\lambda)^{-\frac{1}{2}} du$,

hence, if $t = (\epsilon - u^\lambda)^{\frac{1}{2}} \epsilon^{-\frac{1}{2}}$ is introduced as a new integration variable,

$$\frac{1}{2}\lambda s_0 = \epsilon^{-\frac{1}{2}+1/\lambda} \int_0^1 (1 - \epsilon(1 - t^2))^{\frac{1}{2}} (1 - t^2)^{-1+1/\lambda} dt. \tag{11}$$

In view of $\lambda < 2$, this implies that $s_0 \rightarrow 0$ as $\epsilon \rightarrow 0$.

Ad (ii) It is clear that the arc $\Gamma_{-\epsilon}$, obtained by reflecting $\Gamma_{+\epsilon}$ across the line $u \equiv 0$, is also a geodesic (corresponding to the $-$ sign in (10) at

$s = 0$). Hence, if v_0 is positive and sufficiently small, there exist at least three geodesics joining $(0, 0)$ with $(0, 2v_0)$; namely, $u \equiv 0$ and both $\Gamma_{+\epsilon}$ and $\Gamma_{-\epsilon}$ for suitable $\epsilon > 0$. Actually, there is an infinity of geodesics joining $(0, 0)$ with $(0, 2v_0)$; for example, the "extension" of a geodesic joining $(0, 0)$ with $(0, 2v_0/n)$, where $n = 2, 3, \cdots$, also joins $(0, 0)$ with $(0, 2v_0)$.

Let $v_0 > 0$, and let Γ be an arc joining $(0, 0)$ with $(0, 2v_0)$ and minimizing the arc-length in the sense of (ii). Then, if v_0 is sufficiently small, Γ is a geodesic; cf. the proof of (I). It will be shown that Γ is not a segment of $u \equiv 0$. Consequently, the arc obtained by a reflection of Γ across the line $u \equiv 0$ is also a minimizing arc. Hence, the minimizing arc is not unique; so that (7) shows that the answer to (ii) is in the negative. (Actually, a refinement of the considerations to follow shows that the arcs $\Gamma_{+\epsilon}$ and $\Gamma_{-\epsilon}$, occurring above, are the only minimizing arcs joining $(0, 0)$ with $(0, 2v_0)$, when v_0 is sufficiently small.)

The length (in the metric (7)) of the line-segment joining $(0, 0)$ with $(0, 2v_0)$ is $2v_0$. It follows from (10) that the length of the corresponding sub-arc of $\Gamma_{+\epsilon}$ is $2s_0$. Thus it is sufficient to verify that $s_0 < v_0$ for small ϵ. But (9), for $c = (1 - \epsilon)^{\frac{1}{2}}$, and (10) imply that

$$v_0 = (1 - \epsilon)^{\frac{1}{2}} \int_0^a (1 - u^\lambda)^{-\frac{1}{2}} (\epsilon - u^\lambda)^{-\frac{1}{2}} du, \text{ where } a = \epsilon^{1/\lambda}.$$

If $t = (\epsilon - u^\lambda)^{\frac{1}{2}} \epsilon^{-\frac{1}{2}}$ is introduced as a new integration variable, it follows that

$$\tfrac{1}{2}\lambda s_0 = \epsilon^{-\frac{1}{2}+1/\lambda} (1 - \epsilon)^{\frac{1}{2}} \int_0^1 (1 - \epsilon(1 - t^2))^{-\frac{1}{2}} (1 - t^2)^{-1+1/\lambda} dt.$$

This relation and (11), respectively, imply that if $\frac{1}{2}\lambda\epsilon^{\frac{1}{2}-1/\lambda} v_0$ and $\frac{1}{2}\lambda\epsilon^{\frac{1}{2}-1/\lambda} s_0$ are defined to be their common limiting value when $\epsilon \to 0$, then both of the resulting functions are differentiable at $\epsilon = 0$, having there the derivatives

$$-\tfrac{1}{4}\lambda \int_0^1 (1 - t^2)^{1/\lambda} dt \text{ and } -\tfrac{1}{2} \int_0^1 (1 - t^2)^{1/\lambda} dt,$$

respectively. Since $-\frac{1}{4}\lambda > -\frac{1}{2}$, it follows that the desired inequality, $v_0 > s_0$, is satisfied for sufficiently small $\epsilon > 0$.

Ad (iii). These considerations show that no minimizing arc (in the sense of (ii)) which joins $(0, 0)$ with a nearby point, (u, v), has the direction of the v-axis at $(0, 0)$.

Ad (iv). Cf. the italicized statement at the end of § **5.**

(v) will be settled in § 7.

Ad (vi). As found at the end of the treatment of (ii), there are three geodesics joining $(u, v) = (0, 0)$ with a nearby point $(0, v)$; namely, $\Gamma_{+\epsilon}$, $\Gamma_{-\epsilon}$ and $u \equiv 0$. But the geodesic $u \equiv 0$ did not furnish the minimal arc-length. Hence, in order to conclude that the answer to (vi) is in the negative, it is sufficient to ascertain that the initial point $(0, 0)$ can be replaced by any point of the geodesic $u \equiv 0$. But this is obvious, since, the functions (7) being independent of v, nothing is altered if the (u, v)-plane is arbitrarily translated in the direction of the v-axis.

Remark. In view of the (local) behavior of geodesics on a surface of class C^2 (cf. [5]), it is interesting to note that there exists a surface (of revolution, and admitting of a parametric representation of class C^1) for which (7) is the metric tensor. In fact, if $f(u) = (1 - |u|^\lambda)^{\frac{1}{2}}$, and if $g(u)$ is defined, for small $|u|$, by the quadrature assigned by $(df/du)^2 + (dg/du)^2 = 1$, then a surface of the type just mentioned results by revolving the arc $(y = f(u),\ z = g(u))$ about the z-axis (cf., e. g., [2], p. 67).

7. The second counter-example. In the (u, v)-plane, put

$$(12) \qquad g_{11} = 1 + |v|^\lambda, \quad g_{12} = 0, \quad g_{22} = 1 + |v|^\lambda, \text{ where } 1 < \lambda < 2.$$

These functions are of class C^1. The corresponding equations (2) are

$$(13) \qquad ((1 + |v|^\lambda)u')' = 0 \text{ and}$$
$$v''(1 + |v|^\lambda) - \tfrac{1}{2}\lambda(\operatorname{sgn} v)|v|^{\lambda-1}(u'^2 + v'^2) = 0.$$

Only the geodesics issuing from $(u, v) = (0, 0)$ will be considered.

It is sufficient to deal with the half-plane $v \geqq 0$ (so that the sign of absolute value can be omitted in (13)). The first equation of (13) gives

$$(14) \qquad (1 + v^\lambda)u' = c = \text{const.}$$

The choice $c = 0$ in (14) shows that $u \equiv 0$ is a geodesic. If $c \neq 0$, then $u' \neq 0$ along the geodesic, so that w can be introduced as a parameter on all geodesics, except $u \equiv 0$, issuing from $(0, 0)$.

Since (12) reduces (6) to $(1 + v^\lambda)(v'^2 + c^2(1 + v^\lambda)^{-2}) = 1$, it follows from (14) that either $c = 0$ or

$$(15) \qquad dv/du = \{k(1 + v^\lambda) - 1\}^{\frac{1}{2}}, \text{ where } k = 1/c^2.$$

If $k = 1$, then a solution of (15) satisfying $v(0) = 0$ is not unique. In fact, if $u_0 \geqq 0$ is arbitrary, then

$$v(u) = 0 \text{ if } 0 \leqq u \leqq u_0 \text{ and } v(u) = \{\tfrac{1}{2}(2 - \lambda)(u - u_0)\}^{2/(2-\lambda)} \text{ if } u > u_0$$

represent a solution of (15) when $k = 1$.

Ad (v). In order to show that the answer to (v) is in the negative in the present case, it remains to show that every point (u, v) can be joined with $(0, 0)$ by at most one geodesic arc. It will be sufficient to consider points (u, v) in the quadrant $u \geqq 0, v \geqq 0$.

All geodesics, except $u \equiv 0$, are obtained from (15) by choosing $k \geqq 1$. Geodesics belonging to different values of k $(\geqq 1)$ in (15) have no point $(u, v) \neq (0, 0)$ in common. In order to prove this, suppose the contrary. Then there exist two values of k, say k_1 and k_2 $(< k_1)$, such that the corresponding geodesics, say $v = v_1(u)$ and $v = v_2(u)$, intersect at some $(u^0, v^0) \neq (0, 0)$. Since $v_1'(0) > v_2'(0)$ by (15), it can be supposed that $v_1(u) > v_2(u)$ if $0 < u < u^0$. But then (15) also implies that $v_1'(u^0) > v_2'(u^0)$, since $v_1(u^0) = v_2(u^0) = v^0$. This is a contradiction.

If $k > 1$, then (15) has a (locally) unique solution $v = v(u)$ satisfying $v(0) = 0$. If $k = 1$, it is seen from the last formula line that the set of points common to two solution curves $v = v(u)$ of (15) which satisfy $v(0) = 0$ is either the single point $(u, v) = (0, 0)$ or a segment of the form $(0 \leqq u \leqq u_0, v = 0)$. Furthermore, any geodesic resulting from $k > 1$ has no point $(u, v) \neq (0, 0)$ on the geodesic $u \equiv 0$. Hence, any point (u, v) can be joined with $(0, 0)$ by at most one geodesic.

8. Proof of (III). The assumptions of (III) imply that the Christoffel symbols $\Gamma^i_{jk} = \Gamma^i_{jk}(u, v)$ satisfy a uniform Lipschitz condition in a neighborhood of $(0, 0)$. Hence, any solution of (2) is uniquely determined by its initial conditions. For convenience, it can be supposed that the functions g_{ik} are normalized as follows: $g_{11}(0,0) = g_{22}(0, 0) = 1$ and $g_{12}(0, 0) = 0$. Let (3) denote the geodesic Γ_ϕ determined by the initial conditions $u(0, \phi) = v(0, \phi) = 0$; $u_r(0, \phi) = \cos \phi$, $v_r(0, \phi) = \sin \phi$. Then r is the arc-length (in the mertic (1)) on Γ_ϕ. The functions (3) are defined and continuous on some circle $\mathfrak{C}: (0 \leqq r < a, 0 \leqq \phi < 2\pi)$.

The assumption that the function $g_{ik}(u, v)$ possess partial derivatives satisfying a uniform Lipschitz condition in a (simply connected) neighborhood of $(u, v) = (0, 0)$ has the following consequence: There exists a bounded measurable function $K = K(u, v)$, the "curvature" of (1), for which K

the formula of Gauss-Bonnet is valid. By this is meant that if S is a simple closed curve which is piecewise of class C^2, and T is the interior of S, then

$$(17) \qquad 2\pi - \sum_{j=1}^{n} \alpha_j - \int_S \kappa ds = \iint_T (g_{11}g_{22} - g_{12}^2)^{\frac{1}{2}} K du\, dv,$$

where $\alpha_1, \cdots, \alpha_n$ are the exterior angles (in the metric (1)) at the vertices of S, and κ denotes the geodesic curvature along S. The idea of using (17) as a definition of the Gaussian curvature of (1) is due to Weyl [8], pp. 42-44.

The relation (17) can be verified as follows: Let $g_{ik}^n(u, v)$, where $n = 1, 2, \cdots$, denote three sequences of smooth functions which, on some neighborhood of $(u, v) = (0, 0)$, tend uniformly to the given $g_{ik}(u, v)$, in such a way that, as $n \rightarrow \infty$, the sequences of the partial derivatives $\partial g_{ik}^n / \partial u_j$ and $\partial^2 g_{ik}^n / \partial u_h \partial u_j$ are uniformly convergent and uniformly bounded, respectively. For fixed S (and T), let (17_n) denote the identity which results if the g_{ik} are replaced by g_{ik}^n in (17). Since the left-hand side of (17_n) depends only on S, $\partial g_{ik}^n / \partial u_j$, g_{ik}^n (and $\alpha_j = \alpha_j^n$), the left-hand side of (17_n) tends to that of (17), as $n \rightarrow \infty$. Hence the right-hand side of (17_n), say $\Phi_n(T)$, tends to a limit, say $\Phi(T)$, as $n \rightarrow \infty$. Since each set-function $\Phi_n(T)$ is a Lebesgue integral over T, and since the integrands are uniformly bounded (with respect to u, v and n), it follows that the limit set-function $\Phi(T)$ is a Lebesgue integral, over T, of a bounded function. This proves (17).

9. The proof that $u_r(r, \phi)$, $v_r(r, \phi)$ are of class C^1 on a sufficiently small circle, can be obtained as a modification of the method used in [5]. If $|\phi|$ is sufficiently small, u can be used as parameter on the arc Γ_ϕ, for sufficiently small u. Let $v = v(u; \phi)$, where $0 \leqq u \leqq u_0$, denote this parameterization Γ_ϕ; so that $v(0; \phi) = 0$ and $v_u(0; \phi) = \tan \phi$. As in [5], beginning of § 4, it is sufficient to show that the partial derivative $v_u(u; \phi)$ is of class C^1 in a wedge $(0 \leqq u \leqq u_0, |\phi| < \phi_0)$.

For given ϕ, $\mu > 0$ and $\Delta\phi \neq 0$, let S denote the closed curve consisting of the portion $0 \leqq u \leqq \mu$ of Γ_ϕ; of the segment on the line $u = \mu$ which joins $(u, v) = (\mu, v(\mu; \phi))$ with $(u, v) = (\mu, v(\mu; \phi + \Delta\phi))$; finally, the portion $0 \leqq u \leqq \mu$ of $\Gamma_{\phi+\Delta\phi}$. Then S is piecewise of class C^2, with possible corners only at the points $(0, 0)$, $(\mu, v(\mu, 0))$ and $(\mu, v(\mu, \phi + \Delta\phi))$. At this stage of the proof, it is not evident that the closed curve S is a Jordan curve. Nevertheless, it is clear that the interior T can be defined and (17) can be applied. The result is

$$(18) \qquad \Delta\phi + \pi - \beta - \alpha - \int_{v_1(\mu)}^{v_2(u)} \kappa ds = \int_0^{\mu} \int_{v_1(\mu)}^{v_2(\mu)} (g_{11}g_{22} - g_{12}^2)^{\frac{1}{2}} K\, dv\, du,$$

where $\pi - \Delta\phi$, α, β are the suitably oriented angles (in the metric (1)) at the corners of S, and $v_1(u)$, $v_2(u)$ are abbreviations for $v(u;\phi)$ and $v(u;\phi+\Delta\phi)$, respectively, finally κ denotes geodesic curvature along the line $u=\mu$.

Along the line $u=\mu$, consider a field of parallel vectors (parallel with reference to the metric (1)). If $\theta=\theta(v)$ is a continuous determination of the angle from a vector in this field to the line $u=\mu$, then the integral on the left of (18) is $\Delta\theta=\theta(v_2)-\theta(v_1)$. Also, if $w=w(v)=(w_1(v),w_2(v))$ $\neq(0,0)$ denotes, on the line $u=\mu$, any continuous field of vectors for which $w(v_1)$, $w(v_2)$ are directed along the vectors tangent to Γ_ϕ, $\Gamma_{\phi+\Delta\phi}$, respectively, and if $\omega=\omega(v)$ is a continuous determination of the angle from w to the line $u=\mu$, then $(\pi-\beta)-\alpha=\Delta\omega=\omega(v_2)-\omega(v_1)$. Hence, the left-hand side of (18) is $\Delta\phi+\Delta\psi$, where $\psi=\omega-\theta$ is the angle from the vector w to the corresponding vector in the given field of parallels.

If $w(v)$ is of class C^1, it is easily seen that $\psi=\psi(v)$ is of class C^1. Let

$$w_1(v)\equiv 1 \text{ and } w_2(v)=\{(v_2-v)v_{1u}+(v-v_1)v_{2u}\}/(v_2-v_1),$$

where $v_1=v_1(\mu)$, $v_2=v_2(\mu)$, $v_{1u}=v_{1u}(\mu)$ and $v_{2u}=v_{2u}(\mu)$ (the subscript u denotes differentiation with respect to u). Then $d\psi/dv$ is

$$(g_{11}g_{22}-g_{12})^{\frac{1}{2}}g_{22}^{-\frac{1}{2}}(g_{ik}w_iw_k)^{-1} \text{ times } w_{2v}+\Gamma^1{}_{12}w_2+\Gamma^1{}_{22}w_2{}^2-\Gamma^2{}_{12}-\Gamma^2{}_{22}w_2,$$

where the argument of g_{ik}, $\Gamma^i{}_{jk}$ is (μ,v), and that of w_2, w_{2v} is v (cf., e. g., [7], pp. 131-132, where a similar formula is deduced under the assumption that g_{ik} is the metric tensor of a surface.) Since $w_{2v}=(v_{2u}-v_{1u})/(v_2-v_1)$ $=\Delta v_u/\Delta v$, and $w_2(v)=w_2(v_1)+o(1)$ as $\Delta\phi\to 0$, it follows that there exists a pair of continuous functions $p(\mu,v)>0$, $q(\mu,v)$ satisfying

$$(19)\qquad \Delta\psi=(p(\mu,v_1)+o(1))\Delta v_u+(q(\mu,v_1)+o(1))\Delta v,$$

where $o(1)$ refers to $\Delta\phi\to 0$ and is uniform in μ for $0\leqq\mu\leqq u_0$. This derivation of (19) is valid under the tacit assumption that $v_1(\mu)\neq v_2(\mu)$. It is easily verified that (19) holds also if $v_1(\mu)=v_2(\mu)$.

For the sake of brevity, let $(g_{11}g_{22}-g_{12}{}^2)^{\frac{1}{2}}K$ be denoted by $M=M(u,v)$. Then the boundedness of K implies that

$$\int_{v_1}^{v_2} M(u,v)\,dv=\{M(u,v_1(u))+\delta\}\Delta v(u),$$

where $\delta=\delta(u,\Delta\phi)\to 0$, as $\Delta\phi\to 0$, holds almost everywhere on $0\leqq u\leqq u_0$, and $\delta(u,\Delta\phi)$ is bounded. Hence, (18) can be written as

$$(20)\quad 1+(p+o(1))\Delta v_u/\Delta\phi+(q+o(1))\Delta v/\Delta\phi=\int_0^{\mu}(M+\delta)(\Delta v/\Delta\phi)du.$$

It follows from (20) that the functions $\Delta v(u;\phi)/\Delta\phi$ of u are equicontinuous in ϕ and $\Delta\phi$, for $|\Delta\phi|\neq 0$ sufficiently small. Hence, by the usual arguments occurring in the theory of differential equations, $\Delta v(u;\phi)/\Delta\phi$ has a uniform limit, say $v_\phi(u;\phi)$, as $\Delta\phi\to 0$, and this limit is the unique function $y=y(u)$ satisfying $y(0)=0$ and the linear integro-differential equation

$$1+p(u,v(u;\phi))y'+q(u,v(u;\phi))y=\int_0^{u}M(\mu,v(\mu;\phi))y(\mu)d\mu.$$

This completes the proof that $u_r(r,\phi)$, $v_r(r,\phi)$ are of class C^1 on a sufficiently small circle $\mathcal{C}$.

10. That (3) represents a topological mapping of the sufficiently small circle $\mathcal{C}$ onto a neighborhood of $(u,v)=(0,0)$ and that the Jacobian $\partial(u,v)/\partial(r,\phi)\neq 0$ for $r\neq 0$ can be proved, exactly as in [5], end of § 4, by standard arguments involving Lipschitz's normal coordinates.

It only remains to verify that the element ds^2 of arc-length in the (r,ϕ)-coordinates has the form (4). In what follows, the subscripts r and ϕ, and only these, will denote partial differentiations. Substitution of (3) into (1) shows that the coefficient of dr^2 in ds^2 is

$$(21)\qquad g_{ik}(u(r,\phi),v(r,\phi))u_{ir}u_{kr},\qquad (u_1,u_2)=(u,v).$$

But (21) must be identically 1, since r is arc-length along each of the parameter curves Γ_ϕ, where $\phi=$ const. The coefficient of $2drd\phi$ in ds^2 is

$$(22)\qquad g_{ik}(u(r,\phi),v(r,\phi)u_{ir}u_{k\phi}.$$

Since $u_\phi=v_\phi=0$ holds identically in ϕ at $r=0$, the function (22) vanishes identically if it possesses a partial derivative with respect to r and if this derivative is identically 0.

That (22) has a partial derivative with respect to r follows from the fact that u_r and v_r are of class C^1. This partial derivative is the sum of

$$(23)\qquad (\partial g_{ik}/\partial u_j)(u_{jr}u_{ir}u_{k\phi})+g_{ik}u_{irr}u_{k\phi}$$

and

$$(24)\qquad g_{ik}u_{ir}u_{k\phi r}.$$

Since (21) is identically 1, it follows, by differentiating (21) with respect to ϕ, that the expression (24) reduces to

$$(25) \qquad -\tfrac{1}{2}(\partial g_{ik}/\partial u_j)\, u_{ir} u_{jr} u_{k\phi}.$$

Hence the partial derivative of (22) with respect to r is the sum of (23) and (25). After the addition of (23) and (25), the coefficient of $u_{k\phi}$ is seen to become

$$(26) \qquad g_{ik}u_i'' + (\partial g_{ik}/\partial u_j - \tfrac{1}{2}\partial g_{ij}/\partial u_k) u_i' w_j', \qquad (k = 1, 2),$$

where $' = \partial/\partial r$. It is easily verified that (26) can be written in the form

$$(27) \qquad g_{ik}u_i'' + [ij, k] u_i' u_j', \qquad (k = 1, 2)$$

where $[ij, k]$ is a Christoffel symbol of the first kind. But the expressions (27) vanish by virtue of (2), since (3) represents a geodesic when $\phi = \text{const.}$ Consequently, the partial derivative of (22) with respect to r is 0, and so (22) vanishes identically. This completes the proof of (III).

11. Remark on a non-analytic torse of class C^∞. In the Appendix below, there will be proved a theorem dealing with the situation in which the g_{ik} in (1) are just continuous. The assertion of the theorem will nevertheless be to the effect that, in a certain respect, the behavior of a minimizing arc is locally the same as in the Euclidean case. It is therefore somewhat unexpected that, even if the functions g_{ik} are of class C^∞, a minimizing arc Γ can have the following property: If the Hilbert arc Γ is thought of as an arc in the (u, v)-plane, then the line tangent to Γ at a point P of Γ can intersect Γ at an infinity of points which cluster at P.

In order to see this, let (1) be of the form

$$ds^2 = (1 + f'^2)\,du^2 - 2f'du\,dv + dv^2,$$

where f' denotes the derivative of a smooth function $f = f(u)$ of a single variable. Clearly, this ds^2 can be contracted into $ds^2 = du^2 + (dv - f'du)^2$. Accordingly, $ds^2 = dU^2 + dV^2$, if the parameters U, V are defined by the transformation $U = u$, $V = v - f(u)$, the Jacobian of which is the constant $1 \neq 0$. Hence, the given ds^2 is isometric with the Euclidean ds^2. In particular, the line $V = 0$ in the (U, V)-plane is a geodesic, as well as a minimizing arc, Γ. But this Γ is the curve $v = f(u)$ in the (u, v)-plane. Hence, in order to obtain a Γ of the desired type, it is sufficient to choose $f(u) = \exp(-u^{-2})\sin(u^{-1})$ (if $u \neq 0$, and $f(0) = 0$).

Appendix.

12. It was assumed in all of the above considerations that the coefficient function $g_{ik}(u, v)$ of the positive definite form (1) are of class C', at least. On the other hand, it was pointed out by Hilbert himself ([6], p. 186) that his method leading to the extremal of the arc-length

$$(28) \qquad \left| \int_P^Q ds \right|$$

applies also to (sufficiently small pieces of) surfaces $z = f(x, y)$ for which nothing but the existence of a continuous normal is assumed. Then, the function $f(x, y)$ being of class C', the functions g_{ik} occurring in (1) will be just continuous (perhaps nowhere differentiable), and therefore not in general of class C'. Actually, Hilbert's method supplies for (28) a minimizing arc also when all that is given (on a "geodesically convex" (u, v)-domain) is a set of three continuous functions $g_{ik}(u, v)$ for which the quadratic form (1) becomes positive definite, without being necessarily such as to be embeddable into an (x, y, z)-space.* In what follows, a smoothness property of the Hilbert minimizing curves will be proved which holds under this general assumption.

Let Γ be a rectifiable Jordan arc in the given (u, v)-domain of the Gaussian parameters, and let $P, Q, \cdots$ be points on Γ. Consider the integral

---

* It would be desirable to construct a binary symmetric matrix of the three continuous functions, $g_{ik}(u, v)$, corresponding to which the form (1) is positive definite but such as to fail to become realizable on any surface $X = X(u, v)$ of class C' in the Euclidean space $X = (x, y, z)$. If the orders of continuous differentiability are raised by unity, that is, if the given functions $g_{ik}(u, v)$ are assumed to be of class C' and the non-existence of a surface $X(u, v)$ of class C'' having the line element (1) has to be proved, then an example can be obtained as follows (cf. p. 554 of vol. 72 (1950) of this JOURNAL).:

The functions g_{ik} in (7) are of class C' for $-1 < u < 1$, $-\infty < v < \infty$ (and, as a matter of fact, their partial derivatives of first order satisfy a Hölder condition of index $\lambda - 1$), and have, except on the line $u = 0$, derivatives of arbitrarily high order. An application of the Theorema Egregium to the corresponding line-element (1) shows that, at every point (u, v) at which $0 < |u| < 1$, the Gaussian curvature, K, is

$$K = -(2 - 2|u|^{\lambda})^{-2}\lambda^2 |u|^{2\lambda-2} + (2 - 2|u|^{\lambda})^{-1}\lambda(\lambda - 1)|u|^{\lambda-2},$$

where $1 < \lambda < 2$. Hence, $K = K(u, v)$ does not remain bounded as $(u, v) \to (0, 0)$. In particular, $K(u, v)$ cannot be defined at $(0, 0)$ so as to become continuous. Consequently, (1) cannot possess an embedding of class C'' into the (x, y, z)-space. For, if it did, the representation of K as the quotient of the determinants of the second and first fundamental forms would supply the existence of a continuous $K(u, v)$ at $(u, v) = (0,0)$ also.

(28) along the arc of Γ which is contained between P and Q, with the understanding that the ds in (28) is defined by (1), and let this integral (28) be divided by the integral (28) taken along the chord (that is to say, the rectilinear segment in the Euclidean (u, v)-plane) which joins P with Q. Suppose that this ratio tends to 1 when Q tends to P on Γ (while P is fixed). Then Γ will be said to possess the *Archimedean property* at P. Correspondingly, Γ will be said to have the *uniform Archimedean property* if, as $Q \rightarrow P$, the ratio just mentioned tends to 1 uniformly in P, where P is an arbitrary point of Γ.

In this terminology (cf. [1], Propositions 3-6, pp. 6-10), the result to be proved can be formulated as follows:

(*) *In a sufficiently small (u, v)-domain, let the three functions $g_{ik}(u, v)$ be continuous and such as to make the form* (1) *positive definite, and let Γ be any, sufficiently small, minimizing arc* (*Hilbert*) *of* (28) *contained within the given (u, v)-domain. Then Γ has the uniform Archimedean property.*

We were unable to decide whether or not, under the assumptions of (*) alone, the arc Γ, which always is a rectifiable Jordan arc, must have a tangent at each of its points (or, for that matter, whether or not Γ must be an arc of class C'). Needless to say, the differential equations (2) cannot be written down under the present assumptions, since the Christoffel symbols Γ^i_{jk} are not defined when the functions g_{ik} are not differentiable.

The content of (*) can be illustrated by the following

COROLLARY 1. *Under the assumptions of* (*), *there cannot exist on Γ a point which is a corner in the following sense: Γ has right- and left-hand tangents without having a tangent.*

What is more, this property of lacking corners holds uniformly on Γ. In order to formulate this uniformity, consider the set $\Sigma = \Sigma_\eta$ of line-segments which join a point of the arc PQ with a point of the straight line determined by P and Q, and which form with that line an angle exceeding $\eta > 0$ and not exceeding $\pi - \eta > 0$. Then Corollary 1 of (*) can be refined as follows:

COROLLARY 2. *If the assumptions of* (*) *are satisfied and if $h = h_\eta$ denotes the least upper bound of the lengths of line-segments occurring in Σ_η, then there exists, for every $\epsilon > 0$, a $\delta = \delta_\epsilon$ having the property that*

$$h_\eta < \epsilon \, | PQ | \text{ whenever } | PQ | < \delta \tag{29}$$

holds for any pair of points P, Q on Γ.

In (29), the sign of absolute value denotes distance (for instance, Euclidean distance in the (u, v)-plane).

In the proof of the above statements, the following notation will be used: $\mathrm{e}(PQ)$, $\mathrm{d}(PQ)$, $\mathrm{d}_P(PQ)$ will denote the *lengths of the arc PQ of* Γ in the respective metrics

$$du_i du_i, \quad g_{ik}(u_1, u_2)\, du_i du_k, \quad g_{ik}(u_1{}^0, u_2{}^0)\, du_i du_k,$$

the first being the Euclidean metric in the (u, v)-plane, the second the metric (1), and the third, in which $P = (u_1{}^0, u_2{}^0)$ is fixed, a metric which is affine-equivalent to the first. With references to these three metrics, $\mathrm{e}[PQ]$, $\mathrm{d}[PQ]$, $\mathrm{d}_P[PQ]$ will respectively denote the *lengths of the chord PQ of* Γ.

13. *Proof of* (*). Let $C: u^2 + v^2 \leqq a^2$ be circle on which the functions $g_{ik}(u, v)$ are given as continuous and which contains Γ. Then, on the one hand, there exist positive constants m, M satisfying (5) on C and, on the other hand, there belongs to every $\epsilon > 0$ a $\delta = \delta_\epsilon$ having the property that if P_1, P_2 is any pair of points on C, then $| g_{ik}(P_1) - g_{ik}(P_2) | < \epsilon$ whenever $\mathrm{e}[P_1P_2] < \delta$. In view of Pythagorean inequality $(| a | + | b |)^{\frac{1}{2}} \leqq | a |^{\frac{1}{2}} + | b |^{\frac{1}{2}}$, this implies that, if ξ_1, ξ_2 are arbitrary as in (5), then

$$(30) \qquad | (g_{ik}(P_1)\xi_i\xi_k)^{\frac{1}{2}} - (g_{ik}(P_2)\xi_i\xi_k)^{\frac{1}{2}} | \leqq \epsilon(\xi_1{}^2 + \xi_2{}^2)^{\frac{1}{2}}, \text{ where } \mathrm{e}[P_1P_2] < \delta.$$

Let P, Q be two distinct point of Γ. Then it is clear from the definitions of the arc-lengths introduced at the end of Section 12, that

$$(31) \qquad | \mathrm{d}(PQ) - \mathrm{d}_P(PQ) | \leqq \epsilon \mathrm{e}(PQ),$$

provided that the arc PQ of Γ is within the circle $(u - u^0)^2 + (v - v^0)^2 < \delta^2$, where $P = (u^0, v^0)$. This proviso is satisfied if $\mathrm{d}(PQ) < M\delta$, where M is the constant occurring in (5). Thus it follows from (31) that

$$(32) \qquad \mathrm{d}(PQ) \geqq \mathrm{d}_P(PQ) - \epsilon \mathrm{e}(PQ) \text{ if } \mathrm{d}(PQ) < M\delta.$$

On the other hand,

$$(33) \qquad \mathrm{d}_P(PQ) \geqq \mathrm{d}_P[PQ].$$

In fact, the distance d_P refers to the metric which is affine-equivalent to the Euclidean metric. Hence (33) follows from the fact that the geodesics of such a metric are straight lines.

Finally, (30) implies that, corresponding to (32),

$$\mathrm{d}_P[PQ] \geqq \mathrm{d}[PQ] - \epsilon \mathrm{e}[PQ] \text{ if } \mathrm{d}(PQ) < M\delta. \tag{34}$$

It follows from (32), (33) and (34) that

$$\mathrm{d}(PQ) \geqq \mathrm{d}[PQ] - \epsilon\{\mathrm{e}(PQ) + \mathrm{e}[PQ]\} \text{ if } \mathrm{d}(PQ) < M\delta. \tag{35}$$

On the other hand, $\mathrm{d}[PQ] \geqq \mathrm{d}(PQ)$, since P, Q are on a Hilbert minimizing arc Γ. Hence, from (35),

$$1 \geqq \mathrm{d}(PQ)/\mathrm{d}[PQ] \geqq 1 - \epsilon\{\mathrm{e}(PQ) + \mathrm{e}[PQ]\}/\mathrm{d}[PQ] \text{ if } \mathrm{d}(PQ) < M\delta.$$

Consequently, from (5),

$$1 \geqq \mathrm{d}(PQ)/\mathrm{d}[PQ] \geqq 1 - 2\epsilon/m \text{ if } \mathrm{d}(PQ) < M\delta.$$

Since m, M are independent of ϵ and $\delta = \delta_\epsilon$, the last formula line implies that $\mathrm{d}(PQ)/\mathrm{d}[PQ] \rightarrow 1$, as $Q \rightarrow P$, holds uniformly in P. This proves (*).

14. *Proof of Corollary* 1. Suppose, if possible, that the assertion of Corollary 1 is false. Then Γ has some point P at which there exist a right- and a left-hand tangent, say l_1 and l_2, but no tangent.

After an affine transformation of the (u, v)-plane, it can be assumed that the metric defining d_P is Euclidean. Let γ, where $0 < \gamma < \pi$, be the (Euclidean) angle between l_1 and l_2. With reference to the given P and a sufficiently small $s > 0$, consider that arc QR of Γ which is bisected by P, in the following sense: $\mathrm{d}_P(QP) = \mathrm{d}_P(PR) = s$. If $\psi = \psi(s)$ denotes the (Euclidean) angle between the chords QP and PR, then $\psi(s) \rightarrow \gamma$ as $s \rightarrow 0$. The perpendicular from P to the chord QR divides the angle ψ into two angles $\alpha = \alpha(s)$, $\beta = \beta(s)$, having the sum ψ. Clearly,

$$\mathrm{d}_P[QR] = \mathrm{d}_P[QP] \sin \alpha + \mathrm{d}_P[PR] \sin \beta. \tag{36}$$

On the other hand, it is clear from the proof of (*) that

$$\mathrm{d}_P[QP] \sim \mathrm{d}_P[PR] \sim s \text{ as } s \rightarrow 0, \tag{37}$$

and similarly that

$$\mathrm{d}_P[QR] \sim \mathrm{d}[QR] \text{ as } s \rightarrow 0. \tag{38}$$

It is seen from (36) and (37) that $\mathrm{d}_P[QR] \sim s(\sin \alpha + \sin \beta)$. On the other hand, (38) and (*) imply that $\mathrm{d}_P[QR] \sim 2s$. Consequently, $\sin \alpha + \sin \beta \rightarrow 2$ as $s \rightarrow 0$. But this is impossible, since $\alpha + \beta \rightarrow \gamma$ as $s \rightarrow 0$, where $0 < \gamma < \pi$. This contradiction proves Corollary 1.

Proof of Corollary 2. In the proof of Corollary 1, only the Archimedean property of Γ at a fixed point P was used. If the full force of (*), claiming the uniform Archimedean property of Γ, is applied, the Corollary 2 follows in about the same way as Corollary 1 did.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] Archimedes, *Works* ("On the Sphere and Cylinder," Book I), edited by T. L. Heath, Cambridge, 1897.

[2] L. Bieberbach, *Differentialgeometrie*, Leipzig and Berlin, 1932.

[3] O. Bolza, *Vorlesungen über Variationsrechnung*, Leipzig and Berlin, 1909.

[4] G. Darboux, *Leçons sur la théorie générale des surfaces*, vol. 2, Paris, 1889.

[5] P. Hartman, "On the local uniqueness of geodesics," *American Journal of Mathematics*, vol. 72 (1950), pp. 723-730.

[6] D. Hilbert, Ueber das Dirichlet'sche Prinzip," *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 8 (1900), pp. 184-188.

[7] E. R. van Kampen, "The theorems of Gauss-Bonnet and Stokes," *American Journal of Mathematics*, vol. 60 (1938), pp. 129-138.

[8] H. Weyl, "Ueber die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1916), pp. 40-72.

ON THE ASYMPTOTIC CURVES OF A SURFACE.*

By PHILIP HARTMAN and AUREL WINTNER.

The classical theory of asymptotic curves on surfaces of non-positive Gaussian curvature is based on assumptions of smoothness which are usually left unspecified. On the other hand, if the necessary assumptions are specified *ad hoc,* that is to say, so as to make directly applicable the standard existence and uniqueness theorems needed from the theory of differential equations, then there become involved hypotheses which not only appear to be too severe but at the same time such as to have nothing to do with the geometrical problems at hand.

The purpose of this paper is to dispose of this anomaly. On the one hand, conditions sufficient for the truth of the classical statements will be developed in terms which take into account the specifically geometrical origin of the various differential equations defining the respective problems. On the other hand, examples will show that the conditions to be imposed cannot be omitted.

Both an attraction and a difficulty of the theory is the circumstance that, even if the surface has a parametrization in terms of asymptotic curves, examples show that this parametrization of the surface *can* be less differentiable than some other parametrization of the surface (or, equivalently, less smooth than the parametrization $z = z(x, y)$, where x, y, z are Cartesian coordinates).

Surfaces of negative Gaussian curvature are dealt with in Parts I, II, III. Part IV deals with lines of curvature and applies therefore to surfaces of positive curvature also. Finally, Part V develops the corresponding theory for torses.

PART I. Problems of Uniqueness.

1. Statement of the theorems. Let S be a small piece of a surface of class C'' in the (x, y, z)-space. Thus, if the Gaussian parameters (u^1, u^2) on S are suitably chosen, there exists on S a first fundamental form, $g_{ik}du^i du^k$, which is positive definite, with coefficient functions $g_{ik} = g_{ik}(u^1, u^2)$ which are of class C', and there exists a second fundamental form, $h_{ik}du^i du^k$, in which the coefficient functions $h_{ik} = h_{ik}(u^1, u^2)$ are continuous. The signa-

---

* Received May 28, 1950.

ture of the second fundamental form depends on the Gaussian curvature $K = K(u^1, u^2)$, since $K = \det(h_{ik})/\det(g_{ik})$, where $\det(g_{ik}) > 0$.

Suppose that $K < 0$ on S. This means that the second fundamental form is indefinite throughout. Hence every point (u^1, u^2) determines a distinct pair of directions, $du^1 : du^2$, characterized by the vanishing of the second fundamental form. This pair of directions is called the pair of asymptotic directions (at a given point of S). By an asymptotic curve on S is meant a curve, of class C', along which the direction of the tangent is always an asymptotic direction.

The papers and books known to us either do not specify a precise degree of smoothness required of the surface S, or else they imply that all asymptotic curves together must consist of two locally schlicht families of curves, which can therefore be introduced as parameter curves (if S is small enough). A closer inspection of the questions shows, however, that the situation is much more involved.

The differential equation of the asymptotic curves is

$$h_{ik} du^i du^k = 0, \qquad (i = 1, 2; k = 1, 2), \tag{1}$$

where, by assumption, the functions $h_{ik} = h_{ik}(u^1, u^2)$ have a negative determinant. Hence, (1) can be split into two real differential equations of the form

$$dv/du = f(u, v), \tag{2}$$

where u, v denote u^1, u^2 or u^2, u^1, respectively (the respective functions f are not identical). But if S is just of class C'' (and, correspondingly, the parameters, u and v, on S are so chosen that they supply a C''-parametrization of S), then the functions $h_{ik}(u, v)$ are just continuous, and so nothing like a Lipschitz condition is assured for the continuous function $f(u, v)$ occurring in (2). Hence, it cannot be concluded that (2) has only one solution $v = v(u)$ satisfying an initial condition, say $(v)_{u=0} = 0$.

There arises, however, the question whether the resulting possibility of non-uniqueness can actually occur. It will be proved below that the answer to this question is affirmative:

(i) *If a surface, S, is of class C'' and if the Gaussian curvature is negative at a point, P, of S, then P can issue a continuum of asymptotic curves (each of which is tangent to one of the two distinct asymptotic directions at P).*

This is the more interesting as, under the standard assumptions of differentiability, the asymptotic curves are known to be identical with the characteristics of the partial differential equations defining the problem of deformation (cf., e. g., [2], p. 279), as well as with the characteristics of the partial differential equations on which the classical problem of embedding depends (ibid., p. 252).

The assertion of (i) is anything but obvious. In fact, if $f(u,v)$ is given as an *arbitrary* continuous function corresponding to which the initial condition $(v)_{u=0}=0$ determines more than one solution $v=v(u)$ of (2), then (2) cannot be "embedded" so as to become the differential equation of a family of asymptotic lines (on *some* S). Correspondingly, the proof of (i) will depend on a delicate construction, leading to an $f(u,v)$ for which that "embedding" is possible.

Remark. The claim of (i) becomes false if C'' is replaced by C''' in (i). In fact, the functions h_{ik} then become of class C'; hence, the f in (2) is of class C', and so Lipschitz's uniqueness condition is satisfied.

The negative statement of (i) makes clear the content of the following theorem:

(ii) *If a surface, S, of class C'' is of constant negative curvature, then either of the asymptotic directions at a point, P, of S determines a unique asymptotic curve through P.*

The assertion of (ii) is not made illusory by the Remark preceding (ii). In fact, if S is of class C'' and is given in the form $z=z(x,y)$, where z is a function of class C'', then $K=\det(h_{ik})/\det(g_{ik})$ reduces to

$$K=(rt-s^2)/(1+p^2+q^2)^2, \tag{3}$$

where, as usual, $p=z_x, \cdots, t=z_{yy}$. Since the assumption of (ii) means that $K=K(x,y)$ is a negative constant, say -1, it follows that the surfaces dealt with in (ii) can be thought of represented by those functions $z=z(x,y)$ of class C'' which satisfy the partial differential equation

$$rt-s^2+(p^2+q^2+1)^2=0. \tag{4}$$

But the Monge-Ampère equation (4) (with constant, hence analytic, coefficients) is of hyperbolic type and will therefore undoubtedly have solutions $z=z(x,y)$ which are of class C'' but not smoother.

The proof of (ii) will be such as to allow a generalization to certain surfaces, to be called of class C^{**}, which are not of constant curvature. Let S be called of class C^{**} if it is of class C'' and if it has a C''-parametrization in which the Gaussian curvature satisfies a Lipschitz condition (as a function of the parameters in question). It is clear that, if S is of class C^{**}, the Gaussian curvature satisfies a Lipschitz condition in every C''-parametrization.

Clearly, a surface of class C'' which is of constant curvature must be of class C^{**}. Hence, (ii) is contained in the following theorem:

(iii) *If a surface, S, is of class C^{**} and if the Gaussian curvature is negative at a point, P, of S, then either of the asymptotic directions at P determines a unique asymptotic curve through P.*

Inasmuch as the proof of (i) is somewhat length, (iii) will be proved first.

2. Proof of (iii). There is no loss of generality in assuming that S is given in the form $z = z(x, y)$, where $z(x, y)$ is a function of class C'' on the circle $x^2 + y^2 < 1$, and that the point P in (iii) is $(x, y) = (0, 0)$. Then, since $K = K(x, y)$ is supposed to be negative at $(0, 0)$, it is seen from (3) that, after a suitable rotation of the (x, y)-plane, it can be assumed that

$$t(0,0) \neq 0 \tag{5}$$

$(t = z_{yy})$. On the other hand, since $(u^1, u^2) = (x, y)$, the definition of the second fundamental form shows that (1) reduces to

$$r\,dx^2 + 2s\,dx\,dy + t\,dy^2 = 0. \tag{6}$$

It follows from (3) that (6) can be written in the form

$$s\,dx + t\,dy = \pm(-K)^{\frac{1}{2}}(1 + p^2 + q^2)\,dx. \tag{7}$$

It is understood that s, t, p, q, K in (7) are given functions of (x, y).

In a sufficiently small neighborhood of $(x, y) = (0, 0)$, replace the independent variables x, y by

$$x = x, \qquad q = z_y(x, y) \tag{8}$$

(Legendre). Since $z(x, y)$ is of class C'', the (local) binary mapping $(x, y) \rightarrow (x, q)$ is of class C' and (locally) schlicht. In fact, the Jacobian, $\partial(x, q)/\partial(x, y)$ of (8) is z_{yy}, which, in view of (5), is distinct from 0 at the origin and therefore, by continuity, near the origin.

Accordingly, (7) can be written in the form

$$dq = \pm(-K)^{\frac{1}{2}}(1 + p^2 + q^2)\,dx, \tag{9}$$

where K and p are considered as functions of the independent variables x, q. Hence, if it is ascertained that both of these functions, $K(x; q)$ and $p(x; q)$, satisfy a Lipschitz condition with respect to q, then the proof of (iii) will be complete. In fact, either of the differential equations (9) will then be of the form (2), with $u = x$, $v = q$ and with an f satisfying Lipschitz's condition.

Since the surface S is supposed to be of class C^{**}, both $K = K(x, y)$ and $p = z_x(x, y)$ satisfy a Lipschitz condition in x and y together. On the other hand, the transformation (8) and its inverse are of class C' and preserve, therefore, the existence of a Lipschitz constant (near the origin). Hence, the proof of (iii) is complete.

3. Proof of (i). Let $g(x)$, $h(y)$ denote continuous functions, each a function of a single variable, which satisfy

$$g(0) = 0, \qquad h(0) = 0, \tag{10}$$

and let $G(x)$, $H(y)$ then be defined by the quadratures assigned by $d^2G/dx^2 = g(x)$, $d^2H/dy^2 = h(y)$ and by arbitrary initial conditions. Then

$$z = z(x, y) = G(x) + H(y) + xy \tag{11}$$

represents a surface of class C''. On this surface, the differential equation, (6), of the asymptotic curves reduces to

$$g(x)\,dx^2 + 2dxdy + h(y)\,dy^2 = 0. \tag{12}$$

Hence, it is seen from (10) that the surface (11) has a negative curvature at the point $(x, y) = (0, 0)$, and that the asymptotic directions at this point are those of the coordinate axes. But on an asymptotic curve issuing from $(0, 0)$ in the direction of the x-axis, the variable x can be used as a parameter (for small $|x|$). Then (12) becomes

$$g(x) + 2y' + h(y)y'^2 = 0, \tag{13}$$

where the prime denotes differentiation with respect to x. Hence, (i) is a consequence of the following assertion:

(†) *On x- and y-intervals about $x = 0$ and $y = 0$, respectively, it is possible to define continuous functions, $g(x)$ and $h(y)$, satisfying* (10) *and having the following property: The differential equation* (13) *has more than one solution path $y = y(x)$ passing through the point $(x, y) = (0, 0)$.*

Needless to say, the $h(y)$ *in* (†) cannot satisfy a Lipschitz condition.

Remark. The quadratic equation (13) for y' splits into two differential equations, linear in y', one of which is $y' = f(x, y)$, where

$$f = \{-1 + (1 - gh)^{\frac{1}{2}}\}/h \text{ if } h \neq 0, \text{ and } f = -\tfrac{1}{2}g \text{ if } h = 0.$$

This $f(x,y)$ is continuous (for small x^2+y^2), and it is clear from the connection between (iii) and (†) that it is the above $f(x,y)$ corresponding to which the point $(x,y)=(0,0)$ fails to be a point of uniqueness of the differential equation $y'=f(x,y)$. But the latter clearly is the differential equation of the characteristics of a linear partial differential equation of second order,

$$z_{xx}+2f(x,y)z_{xy}+f^2(x,y)z_{yy}=0. \tag{14}$$

Since $1\cdot f^2-f\cdot f$ vanishes identically, (14) is of parabolic type. If $z=z(x,y)$ denotes the surface S supplied by (iii), it is seen that (14) has a solution $z=z(x,y)$ of class C'' on an (x,y)-domain containing the point $(0,0)$. But the latter issues a continuum of characteristics (in the same direction), namely, those solutions of $y'=f(x,y)$ which satisfy $y(0)=0$. Hence, the classical methods, based on fields of characteristics, cannot lead to any solution $z=z(x,y)$ of (14).

4. Proof of (†). Let $w=w(x)$ be defined by

$$w(x)=-x^6+x^7\sin(x^{-5}) \text{ if } x\neq 0, \text{ and } w(0)=0. \tag{15}$$

Then $w(x)$ is of class C'; in fact, the derivative of (15) is given by

$$w'(x)=-6x^5-5x\cos(x^{-5})+7x^6\sin(x^{-5}) \text{ if } x\neq 0, \text{ and } w'(0)=0. \tag{16}$$

The function $y=w(x)$ satisfies $y(0)=0$, and is a solution of (13) if

$$g(x)=-2w'(x)-h(w(x))w'^2(x). \tag{17}$$

Let the continuous function $g(x)$ be defined by (17) in terms of $w(x)$ and of the continuous function $h(y)$ (the latter is unspecified as yet). The first of the conditions (10) is satisfied by virtue of (16) and (17).

It remains to show that the continuous function $h(y)$ can be so chosen that $h(0)=0$ and that (13) possesses a solution $y=y(x)$ satisfying $y(0)=0$ and $y(x)\not\equiv w(x)$, for small $|x|$. But such an h results by placing

$$h(y)=|2y|^{1/3} \text{ if } y\leqq 0, \text{ and } h(y)=0 \text{ if } y>0. \tag{18}$$

In order to see this, define $y(x)$ by the quadratures assigned by

$$y'(x)=-\tfrac{1}{2}g(x) \text{ if } x\geqq 0 \text{ and } y'(x)=w'(x) \text{ if } x<0 \tag{19}$$

and by the initial condition $y(0)=0$. According to (17), this means that

$$(20)\quad y(x) = w(x) + \int_0^x \tfrac{1}{2}w'^2(t)h(w(t))\,dt \text{ if } x \geqq 0,$$
$$\text{and } y(x) = w(x) \text{ if } x < 0,$$

where h is defined by (18). It will be shown that $y(x)$ is a solution of (13) for small $|x|$, and that $y(x) \not\equiv w(x)$ for small $x > 0$.

The function (15) satisfies $w(x) < 0$ for small $|x| \neq 0$. Hence, it will follow that $y(x) \not\equiv w(x)$ if it is shown that

$$(21)\qquad y(x) > 0 \text{ for small positive } x.$$

On the other hand, since $h(y) = 0$ for $y > 0$, it follows from (13) and the first part of (19), that (20) is a solution of (13) for small $x > 0$ if (21) holds. Thus, in order to complete the proof of (†), it is sufficient to verify (21).

5. Proof of (21). The definition (14) of $w(x)$ shows that $w(x) < -\frac{1}{2}x^6$ for small $|x| \neq 0$. Thus the monotony of the function $h(y)$ implies that $h(w(x)) > h(-\frac{1}{2}x^6) = x^2$. Hence, $y(x) \geqq w(x) + v(x)$ for small $x > 0$, if $v(x)$ is defined by

$$(22)\qquad v(x) = \int_0^x \tfrac{1}{2}t^2 w'^2(t)\,dt.$$

It will be shown that, for some constant $C > 0$,

$$(23)\qquad v(x) > Cx^5 \text{ for small positive } x.$$

The existence of such a C leads, for small positive x, to the inequalities

$$y(x) \geqq w(x) + v(x) > w(x) + Cx^5 > 0.$$

The last inequality follows from $C > 0$ and from (15). Thus, if the existence of a C satisfying (23) is proved, the proof of (21) is complete.

In order to assure the existence of a C, note that (15) shows that, for some constant $c > 0$,

$$(24)\qquad x^2 w'^2(x) \geqq cx^4,$$

provided that $\cos^2(x^{-5}) \geqq \frac{1}{2}$ and that $|x|$ is sufficiently small. These provisos are satisfied if

$$(25)\qquad a_k \leqq x \leqq b_k, \text{ where } a_k^{-5} = (k + \tfrac{1}{4})\pi \text{ and } b_k^{-5} = (k - \tfrac{1}{4})\pi,$$

when k is a large positive integer. If $x > 0$ is sufficiently small and the integer $k = k(x)$ is so chosen that $b_k \leqq x < a_{k-1}$, then $k^{-1} \sim \pi x^5$ as $x \rightarrow +0$. Furthermore, (22) implies that

$$v(x) > \tfrac{1}{2} c \sum_{j=k}^{\infty} \int_{a_j}^{b_j} t^4 dt = c \sum_{j=k}^{\infty} (b_j^5 - a_j^5)/10.$$

But the definitions of a_j and b_j show that

$$\sum_{j=k}^{\infty} (b_j^5 - a_j^5) > \text{const.} \sum_{j=k}^{\infty} j^{-2} > \text{Const.}\ k^{-1},$$

where const. > 0 and Const. > 0. Since $k^{-1} \sim \pi x^5$ as $x \rightarrow +0$, the last two formula lines prove the existence of a constant $C > 0$ satisfying (23).

PART II. Asymptotic Parametrizations.

6. The theorems. According to (iii), every point P of a surface S of class C^{**} of negative Gaussian negative curvature issues exactly two asymptotic curves. Consequently, there exist (locally schlicht) parametrizations of S of the form $X = X(u, v)$, where $X = (x, y, z)$ is a *continuous* vector function of (u, v), and $u =$ const. and $v =$ Const. are asymptotic curves. With a slight modification of the conditions of (iii), this assertion can be improved:

(iv) *Let S be a surface of negative curvature, K, possessing a C''-parametrization in which K is a function of class C' (so that, in particular* (iii) *is applicable). Then, near any point of S, there exist parametrizations $X = X(u, v)$, where $X = (x, y, z)$ is a vector function of class C' in (u, v), and* $u =$ const. *and* $v =$ Const. *are asymptotic curves.*

In particular, if S is a surface of class C'' and of constant negative curvature, then it possesses asymptotic parametrizations of class C'. This assertion seems to be curious, since, starting with a pseudo-sphere of class C'', it supplies a parametrization which is of class C' only. Actually, it turns out that (iv) cannot be improved in this respect:

(v) *Under the assumptions of* (iv), *let S: $X = (x, y, z)$ be (locally) parametrized in the form $X = X(u, v)$, where* $u =$ const. *and* $v =$ Const. *are asymptotic curves. Then no such parametrization of S need be of class C''.*

The assertion (v) remains true even if S is of constant negative curvature. In fact, it will be shown (cf. (viii) below) that if S is of constant negative curvature and if its (Cartesian) parametrization, say $z = z(x, y)$, is not of class C''' near a point, then it *cannot* possess an asymptotic parametrization of class C'' near that point. Thus, (v) is contained in (viii), and so a direct proof of (v) will not be given.

7. Proof of (iv). It can be assumed that S is given in the form $z = z(x, y)$. Then it is clear from the assumptions of (iv) that the function $z(x, y)$ is of class C'' and that $K(x, y)$ is of class C'. Consequently, if (5), (8), (9) are used in the same way as in the proof of (iii), it is seen that the differential equations defining the asymptotic curves reduce to two equations of the form (2), with functions $f(u, v)$ of class C'. Since this implies that the (local) general solution, $v = v(u; u_0, v_0)$, of (2) is a function of class C' in u, u_0 and $v_0 = v(u_0; u_0, v_0)$ together, the assertion of (iv) is now clear from the end of the proof of (iii).

PART III. Asymptotic Tchebychef Nets.

8. The existence theorem. A parametrization $X = X(u, v)$ of a surface S of class C' will be called a Tchebychef parametrization (cf., e. g., [1], § 60-§ 63) if it is of class C' and such as to lead to the following form of the squared line-element on S:

$$|dX(u, v)|^2 = du^2 + 2 \cos \phi \, du \, dv + dv^2, \text{ where } \phi = \phi(u, v). \tag{26}$$

If the subscripts of X refer to partial differentiations, (26) is equivalent to

$$|X_u(u, v)| \equiv 1 \text{ and } |X_v(u, v)| \equiv 1, \qquad (X_u \cdot X_v = \cos \phi). \tag{27}$$

Clearly, $\phi(u, v)$ is a continuous function.

It is a central fact in the differential geometry of pseudo-spheres that the latter possess asymptotic parametrizations satisfying (26) (for instance, Hilbert's theorem on the existence of singularities on pseudo-spheres depends on this fact). A careful perusal of the classical proof reveals however, that the surface is tacitly assumed to be of class C''''. On the other hand, it will turn out that, starting with (iv) alone, it is possible to effect a substantial reduction, $C'''' \rightarrow C''$, of the assumption of smoothness:

(vi) *Let a surface S be of constant negative curvature ($K \equiv -1$) and of class C''. Then there exists, near every point of S, a C'-parametrization, $X = X(u, v)$, which is an asymptotic parametrization satisfying* (27), *i. e.,* (26).

9. Proof of (vi). By the remarks following (iv), the assumptions of (iv) imply, for a vicinity of every point of S, the existence of an asymptotic

C'-parametrization. Let $X = X(u, v)$ be such a parametrization. Then $u =$ const. and $v =$ Const. are asymptotic curves, and the first partial derivatives, X_u and X_v, of the vector function X exist and are continuous. It will be shown that this is possible only if the second mixed partial derivative

$$X_{uv} = X_{vu} \text{ exists and is continuous} \tag{28}$$

and represents a vector parallel to the vector product $[X_u, X_v]$:

$$X_{uv} \cdot X_u = 0, \qquad X_{uv} \cdot X_v = 0. \tag{29}$$

If (28) and (29) are granted at every point (u, v) of S, the proof of (vi) is straightforward. In fact, (28) and (29) imply that $(|X_u|^2)_v = 0$ and $(|X_v|^2)_u = 0$ hold identically. This means that $|X_u|^2$ is a function of u alone, and $|X_v|^2$ a function of v alone, say $|X_u(u, v)| = f(u)$ and $|X_v(u, v)| = g(v)$. Finally, the transition from the latter relations to the assertion, (27), of (vi) requires only a C'-transformation, $(u, v) \rightarrow (u^*, v^*)$, of the trivial form $u^* = u^*(u)$, $v^* = v^*(v)$, that is, a re-parametrization along the asymptotic curves.

Accordingly, only (28) and (29) remain to be proved. Actually, (28) holds if the assumptions of (vi) are generalized to those of (iv):

(vi bis) *If S satisfies the assumptions of* (iv), *then* (28) *holds for every asymptotic C'-parametrization supplied by* (iv).

If (vi bis) is proved, (vi) will follow if (29) is verified under the additional assumption, $K \equiv -1$, of (vi).

10. Proof of (vi bis). If

$$l = (-K)^{-\frac{1}{2}}(1 + p^2 + q^2)^{-1}, \tag{30}$$

then, by (9),

$$x_u = lq_u, \qquad x_v = -lq_v, \tag{31}$$

where $X = X(u, v)$ is the asymptotic C'-parametrization supplied by (iv) and $X = (x, y, z)$. The choice, $+$, made in (30) and (31) for the alternative sign, $\pm$, in (9) is admissible, since u and v can be interchanged.

Clearly, x, q in (9) can be replaced by y, p, respectively. Then what corresponds to (31) is

$$y_u = -lp_u, \qquad y_v = lp_v, \tag{32}$$

the choice of the alternative sign, $\pm$, being committed by (31). In fact, in order to see the effect of the commitment in (31), it is sufficient to write (6) in the form

$$dp\,dx + dq\,dy = 0, \tag{33}$$

and to substitute (32) into (33).

In the (euclidean) plane of the parameters u, v, let R be the positively oriented boundary of a small rectangle, having the vertices (u_1, v_1), (u_2, v_1), (u_2, v_2), (u_1, v_2), where $u_1 < u_2$, $v_1 < v_2$, and put

$$[x] = x(u_1, v_1) - x(u_2, v_1) + x(u_2, v_2) - x(u_1, v_2). \tag{34}$$

Then it is seen from (31) that

$$[x] = -\tfrac{1}{2} \int_R (lq_u\,du + lq_v\,dv). \tag{35}$$

Consider the point transformation $(u, v) \rightarrow (p, q)$ defined by the inverse of

$$p = z_x(x, y), \quad q = z_y(x, y), \text{ where } x = x(u, v), \quad y = y(u, v). \tag{36}$$

Since z is a function of class C'' in x, y, and since the components of $X(u, v)$ are of class C' in u, v, it is clear that the transformation $(u, v) \rightarrow (p, q)$ is of class C'. It has a local inverse of class C', since its Jacobian,

$$J = \partial(p, q)/\partial(u, v), \tag{37}$$

does not vanish. In fact, (37) is identical with the product of $rt - s^2$ and $\partial(x, y)/\partial(u, v)$, where $rt - s^2 < 0$, by (3), and $\partial(x, y)/\partial(u, v) \neq 0$, since $X = X(u, v)$ is an admissible C'-parametrization. Incidentally, there is no loss of generality in assuming that (37) is positive; for, if it is negative, it becomes positive upon the substitution $(u, v) \rightarrow (-u, v)$. But the latter can be admitted, since it amounts to a re-orientation of the asymptotic curves $v =$ Const.

Accordingly, if R is small enough, the substitution $(u, v) \rightarrow (p, q)$ is applicable to (35). This gives

$$[x] = -\tfrac{1}{2} \int_R l\,dq; \text{ hence, } [x] = -\tfrac{1}{2} \int\int_Q l_p\,dp\,dq,$$

if Q denotes the interior of R (Green). Consequently, if p, q are replaced by u, v in the last integral, it follows that

$$[x] = -\tfrac{1}{2}\int_{v_1}^{v_2}\int_{u_1}^{u_2} l_p J\,du\,dv, \tag{38}$$

the Jacobian (37) being positive.

If R is made to shrink to a point (u, v), it is seen from (34) and (38) that the second mixed partial derivative x_{uv} at that point (u, v) exists, is equal to x_{vu} and is represented by

$$x_{uv} = -\tfrac{1}{2} l_p J. \tag{39}$$

In view of (30), this implies that $x_{uv} = x_{vu}$ is continuous. It is proved in the same way the y_{uv} and y_{vu} exist, are equal and continuous and that, corresponding to (39),

$$y_{uv} = -\tfrac{1}{2} l_q J. \tag{40}$$

Hence, in order to complete the proof of (vi bis), only the existence of a continuous z_{uv} remains to be ascertained. But this follows by differentiating, with respect to v, the identity

$$z_u = p x_u + q y_u, \tag{41}$$

in which each of the functions p, q, x_u, y_u possesses a continuous partial derivative with respect to v.

11. Completion of the proof of (vi). What remains is the verification of (29) under the assumption $K \equiv -1$. But the latter reduces (30) to

$$l = (1 + p^2 + q^2)^{-1}. \tag{42}$$

It follows therefore from (39) and (40) that

$$x_{uv} = p l^2 J, \qquad y_{uv} = q l^2 J. \tag{43}$$

In addition,

$$z_{uv} = -l^2 J. \tag{44}$$

In fact, differentiation of (41) with respect to v gives

$$z_{uv} = p x_{uv} + q y_{uv} + p_v x_u + q_v y_u.$$

Hence, if x_{uv}, y_{uv} are substituted from (43), and x_u, y_u from (31), (32),

$$z_{uv} = (p^2 + q^2) l^2 J + (p_v q_u - p_u q_v) l.$$

But this can be contracted into the form (44) if use is made of the definitions, (37) and (42), of J and l.

Let $N = N(x, y)$ denote the normal vector of $S: z = z(x, y)$, oriented in such a way that the direction cosines of N become a scalar multiple of the vector $(p, q, -1)$. Then it is seen from (43) and (44) that, if S is represented in the form $X = X(u, v)$, the vector $X_{uv} = (x_{uv}, y_{uv}, z_{uv})$ is a scalar multiple of N. Since the tangent plane of S is spanned out by the vectors X_u, X_v and is normal to N, it follows that the vector X_{uv} is perpendicular to both vectors X_u, X_v, as claimed by (29).

This completes the proof of all the statements made above.

12. On the smoothness of $\varphi(u, v)$ in (26). The assertion (vi) corresponds to (iv); what corresponds to (v) is the statements that the continuous function $\phi(u, v)$ in (26) need not be of class C'. Actually, it will turn out that $\phi(u, v)$ cannot be of class C' unless S is of class C''' (in some parametrization, say $z = z(x, y)$). In this direction, the following fact will first be proved:

(vii) *Under the assumptions of* (vi), *the continuous function* $\phi = \phi(u, v)$ *in* (26), *belonging to an asymptotic Tchebychef parametrization* $X = X(u, v)$ *of* S, *is of class* C' *if and only if the corresponding parametrization* $X = X(u, v)$ *is of class* C''.

It is clear from (27) that if $X = X(u, v)$ is of class C'', then $\phi(u, v)$ is of class C'. Thus, only the converse needs a proof. In view of (vi bis), this converse is contained in the following lemma:

LEMMA. *Let* S *be a surface of class* C'', *and* $X = X(u, v)$ *a* C'*-parametrization of* S *with the properties that the coefficients* $g_{ik} = g_{ik}(u, v)$ *in*

$$|\,dX(u, v)\,|^2 = g_{11}du^2 + 2g_{12}du\,dv + g_{22}dv^2 \tag{45}$$

are of class C' *and that* (28) *holds. Then the parametrization* $X = X(u, v)$ *of* S *is of class* C''.

This lemma has applications other than (vii), as is indicated by (vi bis) and a similar result ([3], p. 724) concerning geodesic polar coordinates. The proof of the lemma will make it clear that the conclusion $C^n \rightarrow C^{n+1}$ can be extended from $n = 1$ to any n.

13. Proof of the Lemma. It will be shown that $X_{uu}(u, v)$ exists and is continuous. To this end, let

$$\Delta g = \{g(u + h, v) - g(u, v)\}/h,$$

where g is a (scalar or vector) function, $|\,h\,| \neq 0$ is a small number and u, v are fixed, and put

$$A = \tfrac{1}{2}\{X_u(u+h, v) + X_u(u, v)\}, \qquad B = X_v(u, v), \qquad C = N(u, v).$$

If $|h|$ is sufficiently small, the vectors A, B, C are linearly independent. Hence, there exist three (unique) scalars $\alpha = \alpha_h(u, v)$, $\beta = \beta_h(u, v)$, $\gamma = \gamma_h(u, v)$ satisfying

$$(46) \qquad \Delta X_u = \alpha_h A + \beta_h B + \gamma_h C.$$

If (46) is multiplied (scalarly) by A, B, C, respectively, there results the following system of equations for α_h, β_h, γ_h:

$$\begin{aligned} A \cdot \Delta X_u &= \alpha_h\{g_{11}(u, v) + o(1)\} + \beta_h\{g_{12}(u, v) + o(1)\} + o(1)\gamma_h, \\ B \cdot \Delta X_u &= \alpha_h\{g_{11}(u, v) + o(1)\} + \beta_h g_{12}, \\ C \cdot \Delta X_u &= \alpha_h o(1) + \gamma_h, \end{aligned}$$

where the $o(1)$-terms refer to $h \rightarrow 0$. The determinant of this system is $\det(g_{ik}) + o(1)$, which does not vanish for sufficiently small $|h|$. Hence, the last three equations can be solved for α_h, β_h, γ_h. It follows that, as $h \rightarrow 0$, the numbers α_h, β_h, γ_h tend to limits, α_0, β_0, γ_0, which are continuous functions of (u, v), if the same is true of $A \cdot \Delta X_u$, $B \cdot \Delta X_u$, $C \cdot \Delta X_u$.

It will be shown that the latter scalar products tend to continuous limits, as $h \rightarrow 0$. It is seen from the definition of g_{11} that

$$\tfrac{1}{2}\Delta g_{11} = \tfrac{1}{2}\{X_u(u+h, v) + X_u(u, v)\} \cdot \Delta X_u.$$

It follows that, as $h \rightarrow 0$, the scalar product $A \cdot \Delta X_u$ tends to the continuous function $\tfrac{1}{2}\partial g_{11}/\partial u$, since it is assumed that g_{11} is of class C'. Similarly, the identity

$$\Delta g_{12} = X_v(u, v) \cdot \Delta X_u + X_u(u+h, v) \cdot \Delta X_v$$

implies that $B \cdot \Delta X_u$ tends to the continuous function $\partial g_{12}/\partial u - X_u \cdot X_{uv}$, since it is assumed that g_{12} is of class C' and that (28) holds. Finally, the identity $X_u \cdot N \equiv 0$ leads to

$$0 = \Delta(X_u \cdot N) = N(w, v) \cdot \Delta X_u + X_u(u+h, v) \cdot \Delta N.$$

Since S is of class C'' in some parametrization, its unit normal vector, which is invariant up to a factor ± 1, is of class C' in any C'-parametrization of S. Thus, the last identity implies that the scalar product $C \cdot \Delta X_u$ tends, as $h \rightarrow 0$, to the continuous function $-X_u \cdot N_u$. It follows that the limits, α_0, β_0, γ_0, of α_h, β_h, γ_h exist and are continuous. Hence, (46) shows that $X_{uu}(u, v)$ exists and has the value

$$\alpha_0 X_u(u, v) + \beta_0 X_v(u, v) + \gamma_0 N(u, v),$$

and is therefore continuous. It is shown in a similar manner that $X_{vv}(u,v)$ exists and is continuous. This completes the proof of the Lemma.

14. On the smoothness of asymptotic Tchebychef parametrizations. The remark following (v) and that preceding (vii) will now be proved. These remarks are contained in the following assertion:

(viii) *Let a surface S be of constant negative curvature* ($K \equiv -1$) *and suppose that S possesses an asymptotic Tchebychef parametrization $X = X(u,v)$ of class C''. Then S is of class C''' in a Cartesian parametrization, say $z = z(x,y)$.*

Proof of (viii). According to (26), the elements of the first fundamental matrix $(g_{ik}) = (g_{ik}(u,v))$ are given by $g_{11} = g_{22} \equiv 1$ and $g_{12} = \cos\phi$. In particular,

$$\det(g_{ik}) = \sin^2\phi, \text{ hence } \sin\phi \neq 0. \tag{47}$$

The elements of the second fundamental matrix are given by $h_{11} = h_{22} \equiv 0$ and $h_{12} = \pm\sin\phi$. In fact, $h_{11} = h_{22} = 0$ follows from the fact that $u =$ const. and $v =$ Const. on asymptotic curves; while $h_{12}{}^2 = \sin^2\phi$ is equivalent to the identity $-1 \equiv K = \det(h_{ik})/\det(g_{ik})$. (To avoid ambiguity, it can be supposed that $h_{12} = \sin\phi$; otherwise (u,v) can be replaced by the parameters $(-u,v)$, the latter replacement being equivalent to a re-orientation of the asymptotic lines $v =$ Const.)

The derivation formulae of Weingarten for the partial derivative of the unit normal vector $N(u,v)$ supply the relations

$$N_u = (X_u\cos\phi - X_v)/\sin\phi, \qquad N_v = -(X_u - X_v\cos\phi)/\sin\phi,$$

which show that $N(u,v)$ is of class C''.

Let the coordinate system (x,y,z) be so chosen that S has a C''-parametrization of the form $z = z(x,y)$. Since the vector $N(u,v)$ is $\pm(p,q,-1)\{1+p^2+q^2\}^{-\frac{1}{2}}$, where $p = z_x(x,y)$, $q = z_y(x,y)$ are considered as functions of u, v, it follows that p, q are of class C'' as functions of u, v. But the parameter transformation $(x,y) \rightarrow (u,v)$ is of class C'' (with non-vanishing Jacobian). Hence, p, q are of class C'' as functions of x, y also. Consequently, $z = z(x,y)$ is a function of class C'''. This proves (viii).

15. On the Theorema Egregium. If Tchebychef parameters are used on a surface having the Gaussian curvature $K \equiv -1$, then the Theorema Egregium is known to reduce to

$$\phi_{uv} = \sin\phi, \tag{48}$$

or to an integrated form which is the standard formula of Hazzidakis,

$$(49)\qquad [\phi]=\int_{v_1}^{v_2}\int_{u_1}^{u_2}\sin\phi(u,v)\,du\,dv,$$

where

$$(50)\qquad [\phi]=\phi(u_1,v_1)-\phi(u_2,v_1)+\phi(u_2,v_2)-\phi(u_1,v_2),$$

$\phi=\phi(u,v)$ being the function occurring in (26) (cf., e. g., the presentation [1], § 60-§ 63 and § 245). But as mentioned before (vi), the classical theory of pseudo-spheres S tacitly assumes that S has an artificially high degree of smoothness. Correspondingly, the standard proof of (48) or (49) depends on the assumption C''''. It turns out, however, the drastic reduction, $C''''\rightarrow C''$, mentioned before (vi), is possible for the integrated form (49) also, while the reduction $C''''\rightarrow C'''$ is possible for (48):

(ix) *Under the assumptions and in the notations of* (vi), *the function* $\phi=\phi(u,v)$ *occurring in* (26) *satisfies the identity* (49).

Proof of (ix). According to van Kampen's proof [5] of the theorem of Gauss-Bonnet, the latter theorem holds for surfaces of class C'' and for contours consisting of a finite number of arcs of class C''. In addition, it is clear from van Kampen's proof that, on surfaces of class C'', the Jordan contour can be allowed to consist of a finite number of arcs of class C', if the integral of the geodesic curvature over the contour (i. e., the first term is van Kampen's formula (28), p. 134) is replaced by the sum $\Sigma\Delta\theta$. In this sum, extended over all arcs of class C' which constitute the contour, $\Delta\theta$ denotes the oriented variation of a continuous determination of θ along one of the arcs of class C', and $-\theta$ is the oriented angle from the tangent vector of the arc to a vector transported by Levi-Civita's parallelism along the arc (cf. ibid., p. 131).

Let R be the positively oriented boundary of the small rectangle considered before (34). Then if the preceding form of the theorem of Gauss-Bonnet is applied to the Jordan contour R, and if Q denotes the interior of R, it follows that

$$(51)\qquad \Sigma\,\Delta\theta+\iint_Q K\,dS=2\pi-\Sigma\,\alpha,$$

where $\Sigma\,\alpha$ is the sum of the oriented exterior angles, α, at the vertices of the image of R on the surface S.

It is understood that $dS=|\,dS\,|$ in (51) denotes the surface element on S. Hence $dS = (\det g_{ik})^{\frac{1}{2}} du\, dv$, where the square root is positive, and $g_{11} = 1$, $g_{12} = \cos \phi$, $g_{22} = 1$, by (26). Since $K \equiv -1$, it follows that

$$\iint\limits_{Q} K dS = - \int_{v_1}^{v_2}\int_{u_1}^{u_2} |\sin \phi(u, v)|\, du\, dv,$$

Q being the interior of the rectangle R having the vertices (u_i, v_k), where $u_1 < u_2$, $v_1 < v_2$. Thus (51) reduces to

$$\Sigma\, \Delta\theta - \int_{v_1}^{v_2}\int_{u_1}^{u_2} \sin \phi(u, v)\, du\, dv = 2\pi - \Sigma\, \alpha, \tag{52}$$

if it is assumed that $|\sin \phi| = \sin \phi$. This assumption is admissible, since $\sin \phi(u, v)$ is a continuous function which cannot vanish, by (47), and which goes over into $-\sin \phi$ if (u, v) is replaced by $(-u, v)$ (the latter replacement is equivalent to a re-orientation of the asymptotic lines $v =$ Const.).

Next, the orientation of the rectangle R and the definition of the angles α, given after (51), show that

$$\Sigma\, \alpha = (\pi - \phi^1) + \phi^2 + (\pi - \phi^3) + \phi^4, \tag{53}$$

where

$$\phi^1 = \phi(u_1, v_1), \quad \phi^2 = \phi(u_2, v_1), \quad \phi^3 = \phi(u_2, v_2), \quad \phi^4 = \phi(u_1, v_2). \tag{54}$$

In order to evaluate the remaining sum, $\Sigma\Delta\theta$, which occurs in (52), use will be made of the fact that, according to (27), both vectors $X_u = X_u(u, v)$, $X_v = X_v(u, v)$ are of unit length. On the other hand, by (29) or the end of the proof of (vi), the vector X_{uv} is parallel to the normal vector, N, of S. Finally, two sides of the rectangle R are formed by asymptotic curves $u =$ const., and the other two by asymptotic curves $v =$ Const. Consequently, the vectors X_v, X_u represent parallel transportations of themselves along the former and latter two sides of R, respectively. It follows therefore from the definition of the terms of the sum $\Sigma\Delta\theta$, given before (51), that, along the sides of R, the angle θ can be chosen to be $-\phi$ or ϕ according as $v =$ Const. or $u =$ const. Consequently, from (53),

$$\Sigma\Delta\theta = (\phi_1 - \phi_2) + (\phi_3 - \phi_2) + (\phi_3 - \phi_4) + (\phi_1 - \phi_4). \tag{55}$$

Clearly, (55) can be contracted into $\Sigma\Delta\theta = 2[\phi]$, and (53) into $\Sigma\alpha = 2\pi - [\phi]$, where $[\phi]$ is defined by (50). Hence, (52) reduces to (49). This completes the proof of (ix).

PART IV. Lines of Curvature.

16. On lines of curvature ($K \equiv -1$). The following analogue of (ii) will now be proved:

(x) *If a surface, S, of class C'' is of constant negative curvature, then either of the directions of principal curvature at a point, P, of S determines a unique line of curvature through P.*

Clearly, (x) implies that S has a (locally) schlicht parametrization, $X = X(u, v)$, in which $X(u, v)$ is a continuous vector function having the property that $u = \text{const.}$ and $v = \text{Const.}$ represent lines of curvature, and conversely. Actually, it turns out that such a parametrization, $X = X(u, v)$, of S can be chosen not only to be continuous but of class C' as well. In other words, (iv) and (vi) admit of the following analogue:

(xi) *Let a surface S be of constant negative curvature ($K \equiv -1$) and of class C''. Then there exists, near every point of S, a C'-parametrization $X = Y(u, v)$ in which $u = \text{const.}$ and $v = \text{Const.}$ are lines of curvature.*

Furthermore, such a parametrization can be so chosen that the squared line-element on S becomes of the form

$$|dY(u, v)|^2 = g\,du^2 + (1 - g)^{\frac{1}{2}} dv^2, \qquad 0 < g < 1, \tag{56}$$

where the function $g = g(u, v)$ is continuous.

Proof of (x) *and* (xi). If P is a point of a surface S of class C'' and of negative curvature, Dupin's indicatrix shows that the directions of principal curvature bisect the angles between the asymptotic directions. In particular, if $X = X(u, v)$ denotes, near $(u, v) = (0, 0)$, the local C'-parametrization supplied by (vi), and if $Y(u, v)$ is defined by

$$Y(u, v) = X(\tfrac{1}{2}(u + v), \tfrac{1}{2}(u - v)), \tag{57}$$

then, since $X = X(\text{const.}, v)$ and $X = X(u, \text{Const.})$ are asymptotic curves, $X = Y(\text{const.}, v)$ and $X = Y(u, \text{Const.})$ are lines of curvature (and conversely). Hence, those assertions of (xi) which do not involve g and (56) are equivalent to the particular case $K \equiv -1$ of (iv). Correspondingly, (xi) follows from (ii) by virtue of the C'-parametrization used in (57).

17. A uniqueness theorem for lines of curvature. There is little doubt that the possibility of non-uniqueness, established by (i) for the asymptotic curves on surfaces of class C'', remains realizable if the asymptotic curves are replaced by the lines of curvature and, correspondingly, the assumption $K < 0$ by the assumption that S be free of umbilical points; that is, of points at which $h_{11} : h_{12} : h_{22} = g_{11} : g_{12} : g_{22}$. As is well-known, the exclusion of this proportionality is equivalent to the exclusion of the sign of equality in the inequality $H^2 \geqq K$, which is satisfied by the mean curvature, H, and the Gaussian curvature, K, at every point of any surface of class C''.

Clearly, what would formally correspond to (iii) is the following assertion: If S is a surface of class C^{**} containing no umbilical point, then every point of S issues only two lines of curvature (one in each of the directions of principal curvature). Actually, this formal transfer of (iii) to the case of lines of curvature will remain problematic. In fact, what corresponds to (iii) turns out to be a criterion requiring that *both* H and K be subject to Lipschitz's condition (in a C''-parametrization of S), whereas (iii) assumes a Lipschitz condition only for K; cf. the definition of the class C^{**}, given before (iii). In other words, the adaptation of the device used in the proof of (iii) leads only to the following criterion:

(xii) *If $S : z = z(x, y)$ is of class C'' and if both curvatures, $H = H(x, y)$ and $K = K(x, y)$, of S satisfy uniform Lipschitz conditions in x and y together, then every non-umbilical point of S issues a unique pair of lines of curvature.*

18. **Proof of (xii).** First, the exclusion of umbilical points means that $H^2 - K > 0$. On the other hand, the differential equation defining lines of curvature on a surface $z = z(x, y)$ of class C'' can be written in the form

$$d\{p/(1 + p^2 + q^2)^{\frac{1}{2}}\} = \{H \pm (H^2 - K)^{\frac{1}{2}}\}dx, \tag{58}$$

where $p = z_x$, $q = z_y$ and

$$H = \{2pqs - (1 + p^2)t - (1 + q^2)r\}/(1 + p^2 + q^2)^{3/2},$$
$$K = (rt - s^2)/(1 + p^2 + q^2)^2$$

are thought of as expressed as functions of (x, y). In fact, (58) is just a restatement of the equation of Rodrigues ($2H$ being the sum, and K the product, of the principal curvatures).

Put

$$m = p/(1 + p^2 + q^2)^{\frac{1}{2}}. \tag{59}$$

Then it is clear that both equations (58) are of the form

$$dm = f(x, y)\,dx, \tag{60}$$

where $f(x,y)$ satisfies a Lipschitz condition (in x and y together), since the same is true of both $H(x,y)$ and $K(x,y)$. Hence it is clear that the argument used at the end of the proof of (iii) will become applicable to (60) if it is ascertained that the Jacobian of the binary C'-transformation $(x,y)\rightarrow(x,m)$ is distinct from 0.

According to (59), this Jacobian is identical with the partial derivative $\partial\{p/(1+p^2+q^2)^{\frac{1}{2}}\}/\partial y$. The non-vanishing of the latter at a point (x,y) is readily verified if it is assumed that the directions of the coordinate axes x, y do not correspond to the directions of principal curvature at that point (and therefore near that point). Finally, this assumption can always be satisfied by a suitable rotation of the (x,y)-plane. Clearly, this step corresponds to the normalization (5) in the proof of (iii).

PART V. Torses.

19. Asymptotic lines ($K\equiv 0$). Suppose that a surface S of class C'' is a torse, i.e., that $K\equiv 0$. If a point P of S is not a flat point, that is, a point at which the second fundamental matrix (h_{ik}) vanishes, then, since $K\equiv 0$ means that $\det(h_{ik})\equiv 0$, the point P determines a unique asymptotic direction $du_1:du_2$ satisfying (1). There arises the question whether P can issue more than one asymptotic curve. It turns out that, as in the case $K\equiv -1$ treated in (ii), this possibility cannot arise:

(xiii) *If a surface, S, of class C'' is a torse, then every non-flat point, P, of S issues a unique asymptotic curve. The asymptotic curves are straight lines, along which the tangent plane is invariable, and are the generators of the torse S.*

This assertion is a consequence of the considerations in [4], p. 770. Those considerations also show that if S is given in the form $z=z(x,y)$ for small x^2+y^2, and if $P:(x,y)=(0,0)$ is not a flat point and therefore, without loss of generality, $z_{yy}(0,0)\neq 0$, then the asymptotic curves and only these are determined by the equation $q=\text{const.}$, where $q=z_y(x,y)$; cf. (9) above.

20. Asymptotic parametrizations ($K\equiv 0$). Theorem (xiii) and its proof lead to the natural parametrization of torses:

(xiv) *If S is a torse of class C'' containing no flat point, then every sufficiently small piece of S possesses a C'-parametrization of the form*

$$X(u,v)=A(u)v+B(u), \qquad (A(u)\neq 0), \tag{61}$$

where X denotes the vector (x, y, z), and $A(u)$, $B(u)$ are vector functions of class C' satisfying

$$\det(A, A', B') \equiv 0 \qquad (' = d/du). \tag{62}$$

Proof of (xiv). If S is given in the form $z = z(x, y)$, introduce, instead of x, y, the parameters x, q defined by the C'-transformation (8). The latter has a non-vanishing Jacobian near the origin, since (5) can be assumed. Thus the resulting parametrization, say $X = X(q, x)$, of S is of class C'. Furthermore, by (xiii) and the remark following it, all asymptotic curves and only these are given by $q = \text{const.}$ Accordingly, the direction of the vector $\partial X(q, x)/\partial x$ is independent of x. But the first component of $X(q, x)$ is x; hence, the first component of $\partial X(q, x)/\partial x$ is 1. Consequently, $\partial X(q, x)/\partial x$ is independent of x. It follows therefore by a quadrature that $X(q, x)$ appears in the form (61) by placing $u = q$, $v = x$. Furthermore, $A \neq 0$, since $A = \partial X/\partial v$. In addition, $A(u)$ and $B(u)$ are of class C', since $X(u, v)$ is. Finally, (62) is clear from the fact that the normal vector is constant along the asymptotic lines $u = \text{const.}$ (that is, from the fact that the vector product of $A'v + B'$ into A is parallel to the vector product of B' into A). This proves (xiv).

21. On the smoothness of the parametrization (61). Although (xiv) is an analogue of (vi), the assertion of (viii) does not have an analogue. In fact, if $f(x)$ is a function of class C'' for $-1 < x < 1$ and S is the surface given by $z = z(x, y) = f(x)y$, then S is a torse of class C''. It possesses a parametrization (61), where $x = u$, $v = y$ and $A(u) = (1, 0, f(u))$, $B(u) = (0, 0, 0)$. This parametrization of S is of class C'', but is not of class C''' unless $f(x)$ is.

Thus there arises the question as to whether or not (v) has an analogue. The answer is affirmative:

(xv) *Under the assumptions of* (xiv), *let $S: X = (x, y, z)$ be parametrized (locally) in the form* (61). *Then no such parametrization need be of class C''.*

22. Proof of (xv). Let $f(u)$ be a function of class C'' on the interval $-1 \leqq u \leqq 1$, and put $X = X(u, v)$, where, if $X = (x, y, z)$,

$$x = v, \qquad y = -f'(u)v + u \tag{63}$$

and

$$z = \{f(u) - uf'(u)\}v + \tfrac{1}{2}u^2. \tag{64}$$

Clearly, the Jacobian $\partial(x, y)/\partial(u, v)$ is -1 at $(u, v) = (0, 0)$. Hence, in a vicinity of $(u, v) = (0, 0)$, the surface (63)-(64) has an equation of the form $z = z(x, y)$. It will be shown that this surface, S, is of class C'' in its (x, y)-parametrization; that S is a torse; and that the point P corresponding to $(x, y) = (0, 0)$ is not a flat point (as is not, therefore, any point sufficiently close to P).

A simple calculation involving the Jacobian matrix of the transformation (63) shows that

$$z_x = f(u) \text{ and } z_y = u, \tag{65}$$

where $u = w(x, y)$. Since u is a function of class C' in (x, y), and f is a function of class C'' in u, it follows that z_x, z_y are functions of class C' in (x, y). Hence, S is of class C'' in terms of its (x, y)-parametrization.

It follows from (65) that $z = z(x, y)$ satisfies the partial differential equation $rt - s^2 = 0$; so that, by (3), the Gaussian curvature of S vanishes identically. It is also seen from (65) and (63) that z satisfies the normalization (5); so that $(x, y) = (0, 0)$ is not a flat point.

In order to complete the proof of (xv), it remains to show that if $f(u)$ is a suitably chosen function of class C'', then no C''-parametrization of any vicinity of P: $(0, 0)$ on S is of the form (61). It turns out that S must have this property whenever $f(u)$ fails to have a third derivative at $u = 0$.

First, (63)-(64) is a C'-parametrization of the form (61). Let $u = u(\alpha, \beta)$, $v = v(\alpha, \beta)$ be a change of parameters which transforms (61), that is, (63)-(64), into a parametrization of the same form, say

$$X(\alpha; \beta) = L(\alpha)\beta + M(\alpha). \tag{61 bis}$$

Then, since $w = \text{const.}$ and $\alpha = \text{Const.}$ are asymptotic curves, $u(\alpha, \beta)$ is a function of α alone, say $u = u(\alpha)$. For a fixed u or α, the variables v or β are, up to a translation and a change of scale, the arc-lengths along the asymptotic curves. Hence $v(\alpha, \beta)$ is of the form $g(\alpha)\beta + h(\alpha)$, where $g(\alpha) \neq 0$. Consequently, a necessary condition for a transformation $u = u(\alpha, \beta)$, $v = v(\alpha, \beta)$ taking (63)-(64) into the form (61 bis) is the following structure:

$$u = u(\alpha), \qquad v = g(\alpha)\beta + h(\alpha); \qquad\qquad g(\alpha) \neq 0. \tag{66}$$

Suppose, if possible, that (63)-(64) can be transformed into (61 bis) in such a way that $(\alpha, \beta) = (0, 0)$ corresponds to $(u, v) = (0, 0)$ and that the resulting vector function $X(\alpha; \beta)$ in (61 bis) is of class C'' in a vicinity of $(\alpha, \beta) = (0, 0)$. If $X = (x, y, z)$, the first of the equations (63) shows that $x = x(\alpha;\beta)$ is identical with $g(\alpha)\beta + h(\alpha)$. It follows that $g(\alpha)$ and $h(\alpha)$ are of class C'' near $\alpha = 0$. On the other hand, the second of the equation (63) shows that $y = y(\alpha;\beta)$ reduces to

$$-f'(u(\alpha))g(\alpha)\beta - f'(u(\alpha))h(\alpha) + u(\alpha).$$

Hence, $f'(u(\alpha))g(\alpha)$ and $-f'(u(\alpha))h(\alpha) + u(\alpha)$ are of class C'' near $\alpha = 0$. Since $g(\alpha) \neq 0$, it follows that $f'(u(\alpha))$, and therefore $u(\alpha)$, is of class C'' near $\alpha = 0$.

Thus the three functions of α which occur in (66), namely, $u(\alpha)$, $g(\alpha)$ and $h(\alpha)$, are of class C''. In order that (61 bis) be an admissible parametrization of S, it is necessary that the Jacobian $\partial(u,v)/\partial(\alpha,\beta)$ be distinct from 0 at $(\alpha,\beta) = (0,0)$. This Jacobian is $g(\alpha)du(\alpha)/d\alpha$, by (66). But if $f(u)$ fails to have a third derivative at $u = 0$, then, since $du(\alpha)/d\alpha \neq 0$ at $\alpha = 0$, the function $f'(u(\alpha))$ of α cannot have a second derivative at $\alpha = 0$. This completes the proof of (xv).

23. On lines of curvature ($K \equiv 0$). The following analogue of (x) will now be proved:

(xvi) *If S is a torse of class C'', then either of the directions of principal curvature at a non-flat point, P, of S determines a unique line of curvature.*

Proof of (xvi). The simple parametrization of S supplied by (xiv) cannot be inserted into the differential equations of the lines of curvature, since those differential equations assume a parametrization of class C'', whereas (61) is of class C' only. Let, therefore, S be given in a C''-parametriaztion, say as $z = z(x,y)$. Then the lines of curvature are defined by

$$dN + \lambda dX = 0 \tag{67}$$

(Rodrigues), where N denotes the unit normal vector, X the vector (x,y,z), finally λ either of the two principal curvatures.

The assumption $K \equiv 0$ implies that one of the principal curvatures vanishes identically. On the other hand, the assumption that P, and hence every point in a sufficiently small vicinity of P, is a non-flat point ensures that the other principal curvature is distinct from zero. For the first of these two determinations of λ, it is seen from (67) that $dN = 0$, i. e., $N = \text{const.}$ Hence, in the direction of the principal curvature $\lambda = 0$ every point issues exactly one line of curvature, and this line is an asymptotic curve also; cf., e. g., [4], p. 770.

Since the two directions of principal curvature at a given point are orthogonal, the remaining case, $\lambda \neq 0$, of (67) is equivalent to the differential equations of the orthogonal trajectories of the sheaf of the asymptotic curves. But in terms of the parametrization (61) of S, the asymptotic curves are represented by $u = \text{const.}$ Hence, the differential equation of the orthogonal trajectories in question is

$$g_{12}(u, v) + g_{22}(u, v)\,dv/du = 0. \tag{68}$$

It is readily verified from (61) that (68) reduces to

$$(A \cdot A')v + (A \cdot B') + (A \cdot A)\,dv/du = 0, \qquad (' = d/du), \tag{69}$$

where the dot denotes scalar multiplication. Since (69) is a linear differential equation for $v = v(u)$, with continuous functions of u as coefficients, the solutions of (69) are uniquely determined by their initial conditions. This proves (xvi).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] L. Bianchi, *Lezioni di geometria differenziale*, 3rd ed., Bologna, 1922.

[2] G. Darboux, *Leçons sur la théorie générale des surfaces*, vol. 3, Paris, 1894.

[3] P. Hartman, "On the local uniqueness of geodesics," *American Journal of Mathematics*, vol. 72 (1950), pp. 723-730.

[4] ——— and A. Wintner, "On the fundamental equations of differential geometry," *ibid.*, pp. 757-774.

[5] E. R. van Kampen, "The theorems of Gauss-Bonnet and Stokes, *ibid.*, vol. 60 (1938), pp. 129-138.

ON THE SMALL DIVISORS IN INTEGRATIONS BY LAPLACE TRANSFORMS.*

By AUREL WINTNER.

1. If α is a matrix, of m rows and m columns, let $|\alpha|$ denote the greatest of the absolute values of its m^2 elements. Correspondingly, if $\alpha = \alpha(s)$, where $0 \leqq s < \infty$, let

$$[\alpha] = \int_0^\infty |d\alpha(s)| \tag{1}$$

be the greatest of the total variations of the m^2 scalar functions which constitute the matrix function $\alpha(t)$. Thus the matrix

$$A(t) = \int_0^\infty e^{-ts} d\alpha(s) \tag{2}$$

consists of m^2 Laplace integrals which are absolutely-uniformly convergent for $0 \leqq t < \infty$ if $[\alpha] < \infty$. If in addition

$$\int_0^\infty |d\alpha(s)|/s < \infty, \tag{3}$$

then it is clear from (2) that

$$\int_0^\infty |A(t)|\, dt < \infty. \tag{4}$$

Needless to say,

$$\alpha(+0) = \alpha(0) \tag{5}$$

is necessary for (3). The following theorem will be proved:

In the linear system of differential equations

$$x' = A(t)x, \qquad (' = d/dt), \tag{6}$$

where x is a vector $(x_1, \cdot \cdot \cdot, x_m)$, let the coefficient matrix $A(t)$ be representable, for $0 \leqq t < \infty$, in the form (2), where $[\alpha] < \infty$. Suppose further

---

* Received June 26, 1950.

that α satisfies (3). Then (6) has m linearly independent solution vectors $x = x^1(t), \cdots, x = x^m(t)$ which, when ordered in the form of a matrix $X = (x^1, \cdots, x^m)$, are representable, for $0 \leqq t < \infty$, in the form

$$(7)\quad X(t) = \int_0^\infty e^{-st}\, d\beta(s), \quad \text{where } [\beta] < \infty \text{ and } \beta(+0) - \beta(0) = E$$

(E denotes the unit matrix); so that, by (7),

$$(8)\qquad X(t) = E + X_0(t), \text{ where } X_0(t) \rightarrow 0$$

as $t \rightarrow \infty$, since

$$(9)\qquad X_0(t) = \int_{+0}^\infty e^{-ts}\, d\beta(s).$$

2. Condition (3) is automatically satisfied if the range of integration in (2) does not reach down to $s = 0$, that is, if $\alpha(s) = \text{const.}$ for small non-negative s. For this particular case, the theorem was proved in [2]. The method applied there fails if the points of non-constancy of $\alpha(s)$ are allowed to reach down to $s = 0$, since then the successive approximations lead to "small divisors," substantially of the same type as those occuring in celestial mechanics. Correspondingly, the content of the above theorem is to the effect that the "small divisors" are harmless as long as the "frequency function," $\alpha(s)$, of (2) is subject to the restriction (3).

It turns out that (3) is the best possible condition of its type. In order to see this, suppose that $m = 1$, i. e., that $A(t)$, $\alpha(s)$, $x = X$ are scalars, and let $\alpha(s)$ be non-decreasing and bounded for $0 \leqq s < \infty$. Then $[\alpha] < \infty$ is satisfied, and (2) shows that the general solution of (6) is a constant multiple of

$$x(t) = \exp \int_0^\infty (1 - e^{-st})/s\, d\alpha(s).$$

Clearly, this $x(t)$ tends, as $t \rightarrow \infty$, to ∞ whenever the case $|\, d\alpha(s)\,| = d\alpha(s)$ of the assumption (3) is violated. Hence, nothing like (7) can hold, in general, unless (3) is assumed.

3. The proof of the theorem announced in §1 proceeds as follows:

Each of the m columns of a matrix $X(t)$ is a solution vector $x(t)$ of (6) if

$$(10)\qquad X'(t) = A(t)X(t),$$

and these m solutions will be linearly independent if $\det X(t) \neq 0$ holds for some t (in which case it is known to hold for every t). But if the corollary (8) of (7) is satisfied, then $\det X(t) \neq 0$ for large t (hence, for every t). On the other hand, (7) reduces (10) to

$$X(t) = E - \int_t^\infty A(u)X(u)\,du. \tag{11}$$

Hence, what is sought for is a solution $X(t)$ of (11) which is of the form (7).

To this end, try successive approximations, by placing

$$X(t) = E + \sum_{n=1}^\infty X_n(t), \tag{12}$$

where, in accordance with (11), the terms $X_n(t)$ are assigned by the recursion formula

$$X_{n+1}(t) = -\int_t^\infty A(u)X_n(u)\,du, \tag{13}$$

with

$$X_1(t) = -\int_t^\infty A(u)\,du. \tag{14}$$

In order that (12) should become of the form (7), it is natural to assume that

$$X_n(t) = \int_0^\infty e^{-ts}d\beta_n(s), \tag{15}$$

where $[\beta_n] < \infty$ and

$$\beta_n(+0) = \beta_n(0) \qquad (\text{and, e. g., } \beta_n(0) = 0). \tag{16}$$

On the other hand, direct substitutions show (cf. [2]) that, by virtue of (15) and (2), the relation (14) is formally identical with

$$-s\,d\beta_1(s) = d\alpha(s), \qquad (0 < s < \infty), \tag{17}$$

and the recursion formula (13) with

$$-s\,d\beta_{n+1}(s) = d\alpha(s) * \beta_n(s), \qquad (0 < s < \infty), \tag{18}$$

where the asterisk denotes the operation of convolution (cf. [2]).

It will be proved that if both integrals (1), (3) are finite, then (17), (18) and (16) define, on the closed half-line $0 \leqq s < \infty$, a sequence of matrix functions $\beta_n(s)$ each of which is of finite total variation, $[\beta_n] < \infty$, and that the sum of these total variations is finite,

$$(19) \qquad \sum_{n=1}^{\infty} [\beta_n] < \infty.$$

If this is proved, it will follow that (7) and (11) become satisfied by placing

$$(20) \qquad \beta(s) = \sum_{n=1}^{\infty} \beta_n(s) \text{ if } 0 < s < \infty,$$

and $\beta(0) = \beta(+0) - E$. In fact, the deduction of (7) and (11) will then be the same as in [2] and will, therefore, be omitted.

4. If $\alpha_{ik}(s)$, where $i = 1, \cdots, m$ and $k = 1, \cdots, m$, denote the elements of the matrix $\alpha(s)$, put

$$(21) \qquad \lambda(s) = m \max \int_0^s |\, d\alpha_{ik}(u) \,|, \qquad (0 \leqq s < \infty),$$

where the symbol max refers to i and k together. Thus $\lambda(s)$ is a non-decreasing scalar function. Hence it is clear that the scalar recursion formula

$$(22) \qquad s d\mu_{n+1}(s) = d\{\lambda(s) * \mu_n(s)\}, \text{ where } 0 < s < \infty,$$

will be a majorant of the matrix recursion formula (18) if, corresponding to (17),

$$(23) \qquad s d\mu_1(s) = d\lambda(s), \text{ where } 0 < s < \infty.$$

In fact, since $d\lambda(s) \geqq 0$, it is clear from (23) and (22) that $d\mu_n(s) \geqq 0$, if $\mu_n(s)$ can be formed at all. Hence, in order to prove the existence of the matrix function $\beta_n(s)$ and the convergence of the series (19), it is sufficient to show that (23) and (22) lead to (non-decreasing) scalar functions $\mu_n(s)$ satisfying

$$(24) \qquad \sum_{n=1}^{\infty} [\mu_n] < \infty, \text{ where } [\mu_n] = \int_0^{\infty} d\mu_n(s) \qquad (d\mu_n(s) \geqq 0).$$

The assumptions to be used are

$$(25) \qquad d\lambda(s) \geqq 0; \quad \lambda(0) = 0, \qquad \lambda(\infty) < \infty$$

and

$$(26) \qquad \int_0^{\infty} s^{-1} d\lambda(s) < \infty$$

(hence $\lambda(+0)=\lambda(0)$). In fact, it is clear from (21) that (25), (26) are equivalent to the convergence of the respective integrals (1), (3).

The proof will be based, not directly on (22) and (23), but on the circumstances that

$$y_{n+1}(t)=\int_t^\infty a(s)y_n(s)\,ds \tag{27}$$

and

$$y_1(t)=\int_t^\infty a(u)\,du \tag{28}$$

correspond to (22) and (23) in the same way as (13) and (14) correspond to (18) and (17) if, corresponding to (2) and (15),

$$a(t)=\int_0^\infty e^{-ts}d\lambda(s) \tag{29}$$

and

$$y_n(t)=\int_0^\infty e^{-ts}d\mu_n(s). \tag{30}$$

Then

$$y(t)=1+\sum_{n=1}^{\infty}y_n(t) \tag{31}$$

corresponds to (12), and

$$y(t)=1+\int_t^\infty a(u)y(u)\,du \tag{32}$$

to (11).

Accordingly, (27), (28) and (31) represent the process of successive approximations to a solution of (32), where $a(t)$ is a given scalar function, (29). What is at stake is an appropriate convergence of this process of successive approximations.

5. First, it will be shown that the functions $\mu_1(s),\mu_2(s)\cdots$ can be defined, for $0\leqq s<\infty$ and in accordance with (30) and (27)-(28), in such a way that

$$d\mu_n(s)\geqq 0;\qquad \mu_n(0)=0,\ \mu_n(\infty)<\infty. \tag{33}$$

In view of (30), this will prove that all functions $y_n(t)$ are defined for $0\leqq t<\infty$.

Insertion of (29) into (28) gives

$$y_1(t) = \int_t^\infty \int_0^\infty e^{-us} d\lambda(s)\, du = \int_0^\infty e^{-ts} s^{-1} d\lambda(s),$$

the last rearrangement being justified by (25) and (26). Accordingly, (30) holds for $n = 1$ if $d\mu_1(s) = s^{-1} d\lambda(s)$. In view of (25) and (26), this means that there exists a $\mu_n(s)$ satisfying the case $n = 1$ of (33). Furthermore, by the parenthetical remark made after (26),

$$\mu_n(+0) = \mu_n(0) \tag{34}$$

holds for $n = 1$.

Suppose that $\mu_n(s)$ has been defined, for a fixed n, in such a way that (33) and (34) are satisfied. Then (30) defines for $0 \leqq t < \infty$ a function $y_n(t)$ which is continuous and bounded (even if (34) is not used). Hence

$$\int_0^\infty | a(u) y_n(u) | \, du < \infty \tag{35}$$

provided that $\int^\infty | a(u) | \, du < \infty$. But (29), where $d\lambda(s) \geqq 0$, shows that $a(t)$ is non-negative. Hence the proviso, on which (35) is based, follows from the convergence of (28), which was verified above.

Thus (35) is true. Hence, (27) defines, for $0 \leqq t < \infty$, a function $y_{n+1}(t)$. The explicit form $y_{n+1}(t)$ follows by inserting (29) and (30) into (27). This gives

$$y_{n+1}(t) = \int_t^\infty \int_0^\infty e^{-us} d_s\{ \int_0^\infty \lambda(s - v) \, d\mu_n(v) \} du. \tag{36}$$

In fact, the rearrangement leading to (36) is justified by Fubini's theorem, since (35) is satisfied and $d\lambda \geqq 0$, $d\mu_n \geqq 0$. Finally, one more application of Fubini's theorem shows that (36) can be rearranged into

$$y_{n+1}(t) = \int_0^\infty e^{-ts} d\mu_{n+1}(s), \tag{37}$$

with

$$d\mu_{n+1}(s) = s^{-1} d_s\{ \int_0^\infty \lambda(s - v) \, du_n(v) \}, \tag{38}$$

where the convergence of the integrals (37) and (38), for $0 \leqq t < \infty$ and $0 < s < \infty$ respectively, is part of the statement.

The representation (37) of $y_{n+1}(t)$ means that (30) holds if n is replaced by $n+1$. Since (37) converges at $t=0$, it also follows that

$$(39) \qquad \int_0^\infty d\mu_{n+1}(s) < \infty, \text{ where } d\mu_{n+1}(s) \geqq 0$$

(the second assertion of (39) follows from (38), where $d\lambda \geqq 0$ and $d\mu_n \geqq 0$). If a suitable constant is added to $\mu_{n+1}(s)$, it is seen from (39) that (33) remains true if n is replaced by $n+1$. Finally, (39) and the presence of the factor s^{-1} in (38) imply that (34) must hold when n is replaced by $n+1$.

This completes the induction from n to $n+1$. Its result can be summarized as follows: Starting with (28), the recursion formula (27) defines a sequence of functions $y_n(t)$ which are representable, for $0 \leqq t < \infty$ and $n = 1, 2, \cdots$, in the form (30), where the functions $\mu_n(s)$ satisfy conditions (22) and (34).

What remains to be proved is that the functions $\mu_n(s)$, thus defined, are such as satisfy (24). This will complete the proof of the theorem announced in §**1**.

Since $d\mu_n(s) \geqq 0$, it is clear from (30) that the assertion is equivalent to the absolute convergence of the series (31) at the point $t=0$. But $a(t) \geqq 0$, by (29), where $d\lambda(s) \geqq 0$. Hence it is seen from (27) and (28) that

$$(40) \qquad | y_n(t) | = y_n(t) \leqq y_n(0), \qquad 0 \leqq t < \infty,$$

and it is also clear that

$$(41) \qquad y_{n+1}'(t) = -y_n(t)a(t), \qquad y_{n+1}(\infty) = 0,$$

where $n \geqq 0$ and

$$(42) \qquad y_0(t) \equiv 1.$$

But an induction leads from (42) and (41) to

$$(43) \qquad y_n(t) = (\int_t^\infty a(s)ds)^n/n!.$$

Since application of (43) at $t=0$ shows that $\Sigma y_n(0) < \infty$, it follows from the remark made before (40) that the proof is complete.

6. It is clear from the above proof of the theorem announced in §**1**, that the addition rule of "spectra," used in [2], holds in the present case also, and that the same is true of "point spectra"; cf. [1]. Thus, if the theorem of §**1** is applied to the case of an $\alpha(s)$ which is a step-function, say

$$\alpha(s) = \sum_{\gamma_n < s} A_n, \text{ where } \gamma_n > 0, \tag{44}$$

then there results the following corollary:

Let $A(t)$ be a matrix function of the form

$$A(t) = \sum_{n=1}^{\infty} A_n e^{-\gamma_n t}, \qquad 0 \leqq t < \infty, \tag{45}$$

where every γ_n is positive, every A_n is a constant matrix, and

$$\sum_{n=1}^{\infty} | A_n | < \infty, \qquad \sum_{n=1}^{\infty} | A_n | / \gamma_n < \infty. \tag{46}$$

Without loss of generality, suppose that, by an insertion of amplitude matrices $A_j, A_k, \cdots$ all of which are the zero matrix, the notation in (44), (45) is chosen as follows: If l and $h_1, h_2, \cdots, h_l$ are arbitrary positive integers, then the sum $h_1\gamma_1 + \cdots + h_l\gamma_l$ occurs as a γ_n. Then (6) has a complete set of linearly independent solution vectors $x(t)$ which, when written as the consecutive columns of a matrix $X(t)$, are representable in the form

$$X(t) = E + \sum_{n=1}^{\infty} B_n e^{-\gamma_n t}, \qquad 0 \leqq t < \infty, \tag{47}$$

where the B_n are constant matrices satisfying

$$\sum_{n=1}^{\infty} | B_n | < \infty, \tag{48}$$

and E denotes the unit matrix.

In fact, (47) and (48) represent (7), while (46) represents (1) and (3) together, since (2) reduces to (45) by virtue of (44).

It may finally be mentioned that, since the proof is based on majorizations, everything remains true if the real half-line $t \geqq 0$ is replaced by the complex half-plane $\Re t \geqq 0$. In particular, the real half-line $t \geqq 0$ can be replaced by the imaginary axis, $\Re t = 0$. Accordingly, the above theorem remains true if, both in (45) and (47), one reads it and $-\infty < t < \infty$ instead of t and $0 \leqq t < \infty$, respectively.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. Wintner, "On the addition of independent distributions," *American Journal of Mathematics*, vol. 56 (1934), pp. 8-16.

[2] ———, "On the classical existence theorem of ordinary differential equations," *ibid.*, vol. 71 (1949), pp. 331-338.

ON AN INEQUALITY OF GOLUSIN.*

By JAMES A. JENKINS.[1]

1. Recently Golusin [1], [2] has proved a number of interesting inequalities for univalent functions of which one, in particular, has proved very useful in obtaining bounds for the coefficients of functions univalent in the unit circle. This inequality states that for a function $f(z)$ regular and univalent in $|z| < 1$ and normalized at $z = 0$ by $f(0) = 0$, $f'(0) = 1$,

$$(1) \qquad |f(z_1) - f(z_2)|/|z_1 - z_2| \geqq |f(z_1)f(z_2)|(1 - r^2)/r^2,$$

where $|z_1| = |z_2| = r$, $0 < r < 1$, with equality holding only for $f(z) = z/(1 + ae^{i\theta}z + e^{2i\theta}z^2)$, $\theta = \frac{1}{2}(\arg z_1 + \arg z_2)$, $\quad -2 \leqq a \leqq 2$.

The first object of this note is to show that this result is an immediate consequence of a result proved some years ago by Grötzsch [4]. This latter result can, in its turn, easily be proved by the general methods of extremal length or quadratic differentials but these proofs are not essentially different from Grötzsch's original proof by his method of strips.

As Golusin did, we shall derive first an inequality from which (1) follows at once. Let

$$(2) \qquad F(\zeta) = \zeta + a_0 + a_1/\zeta + \cdots$$

be meromorphic and univalent for $|\zeta| > 1$, then we shall show

$$(3) \qquad |F(\zeta_1) - F(\zeta_2)|/|\zeta_1 - \zeta_2| \geqq 1 - 1/r^2,$$

where $|\zeta_1| = |\zeta_2| = r$, $r > 1$.

The result of Grötzsch in question is as follows: Let D be a domain of the ζ-plane containing the point at infinity and of finite connectivity, let ζ_1, ζ_2 be finite interior points of D, then $|F(\zeta_1) - F(\zeta_2)|$ attains its minimum within the class of functions $F(\zeta)$ meromorphic and univalent in D and normalized at infinity by (2) uniquely for the function mapping D on the plane slit along arcs of the confocal family of hyperbolae having the images of ζ_1 and ζ_2 as foci, this being determined up to an arbitrary choice of a_0 in (2):

---

* Received June 23, 1950; revised November 7, 1950.

[1] Frank B. Jewett Fellow.

Now let D be the domain $|\zeta| > 1$. For the proof of (3) it is clearly enough to take $\zeta_1 = re^{i\phi}$, $\zeta_2 = re^{-i\phi}$; $1 < r$, $0 < \phi < \pi$. The function $F^*(\zeta) = \zeta + 1/\zeta$ gives a mapping onto a configuration of the required extremal type. Since

$$|F(\zeta_1) - F(\zeta_2)| = 2(r - 1/r)\sin\phi$$

holds for this $F = F^*$, it follows that

$$|F(\zeta_1) - F(\zeta_2)| \geqq 2(r - 1/r)\sin\phi$$

holds for any F. Dividing by $|\zeta_1 - \zeta_2| = 2r\sin\phi$, we obtain (3).

Since the resulting extremal function does not assume the value 0 the result holds also in the more restricted class of functions for which this is true. Other extremal functions in this class are $F^*(\zeta) + a$, $-2 \leqq a \leqq 2$. From this we obtain at once inequality (1) and the indicated extremal functions. It should be remarked that this proof does not require the general result of the existence of a mapping on a canonical hyperbolic slit domain. In the limit with $\zeta_1 = \zeta_2$ the analogous result as above holds, the situation being even more simple. It should be pointed out that Golusin obtained a slightly more general result, but (3) is the form which has been used in all applications so far. Also let us note that other of Golusin's inequalities, notably (17) in [1], can be obtained in a similar simple manner.

2. Now we would like to show how Golusin's inequality (1) can be used to simplify the proof of a result of Szegö [10]: Let the function $f(z) = c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n + \cdots$ be regular and univalent in $|z| < 1$. Then all the sections $c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n$ $(n = 1, 2, \cdots)$ are univalent in the circle $|z| < 1/4$.

Szegö's proof uses a lower bound for the left hand side of the inequality (1) obtained from the distortion theorem. The general method works for $n \geqq 5$ with $n = 2, 3, 4$ requiring special devices, the case $n = 2$ being extremely simple. By the use of inequality (1) the case $n = 4$ is brought under the general method but it seems the case $n = 3$ must be treated by the method which Szegö used, namely by the use of Löwner's expressions for the coefficients c_n; [10], p. 196.

If $c_1 = 1$, then $|c_n| < en$, $|c_5| < 8$; cf. [10].

To prove the result for $n \geqq 4$ it is enough to prove the inequality

$$(4) \qquad |f(z_1) - f(z_2)|/|z_1 - z_2| > \left|\sum_{\nu=n+1}^{\infty} c_\nu (z_1{}^\nu - z_2{}^\nu)/(z_1 - z_2)\right|$$

for $|z_1| = |z_2| = 1/4$, $n \geqq 4$. By (1) and the distortion theorem, the quotient on the left of this inequality is at least

$$(1 - r^2)/r^2 \cdot r^2/(1 + r)^4 = (1 - r)/(1 + r)^3 = 48/125 \qquad (r = 1/4),$$

whereas the expression on the right does not exceed (cf. [10])

$$\frac{40}{256} + \frac{9.25 + 24.5 + 20}{27.256} e < \frac{40}{256} + \frac{41}{256} < \frac{48}{125}.$$

This completes the proof.

3. The problem of univalency of the sections of $f(z)$ was taken up by V. Levin [8] who showed that the n-th section was univalent for $|z| < 1 - 6 \log n/n$ $(n \geqq 17)$, an improvement of Szegö's result when n is large. The preceding approach enables us to give a considerably better estimate. We may take $f(z) = z + c_2 z^2 + \cdots$, $f_n(z) = z + c_2 z^2 + \cdots + c_n z^n$. Again we use the fact that $f_n(z)$ will be univalent for $|z| \leqq r_n$ provided inequality (4) holds with $|z_1| = |z_2| = r_n$. This surely will be the case if the right hand side of this inequality is less than $(1 - r_n)(1 + r_n)^{-3}$. For this right hand side we have the upper estimate $\sum_{\nu=n+1}^{\infty} |c_\nu| \nu r_n^{\nu-1}$. To sharpen our final result somewhat we will now use Golusin's estimate $|c_\nu| < \frac{3}{4}e$, [3]: We then obtain above the further bound,

$$\tfrac{3}{4} e r_n^n (1 - r_n)^{-3} \{n^2 (1 - r_n)^2 - (2n - 1) r_n + 2n + 1\}.$$

Thus we will be assured of the desired inequality provided

$$r_n^n (1 - r_n)^{-4} \{n^2 (1 - r_n)^2 - (2n - 1) r_n + 2n + 1\} < (6e)^{-1}.$$

Substituting $r_n = 1 - \alpha/n$ and reducing, we see that this will be the case provided

$$n^4 e^{-\alpha} (\alpha^2 + 2\alpha + 2) \alpha^{-6} < (6e)^{-1}.$$

In order that this be satisfied we may take $\alpha = \beta_n \equiv 4 \log n - \log(4 \log n)$ if n is sufficiently large.

The result is then that the n-th section is univalent for $|z| \leqq 1 - \beta_n/n$. Our result is meaningful only if the right hand side is positive, i. e., for n large enough. Simple calculations show that both conditions on n are satisfied provided $n \geqq 8$. For large n the dominating term in our estimate is $1 - 4 \log n/n$, replacing Levin's $1 - 6 \log n/n$. Let us remark that this could not be improved beyond $1 - 2 \log n/n$, as was shown by Robertson [9].

4. The method of § **2** can, in addition, be used to establish a conjecture of Joh [6]. The result we shall prove is as follows:

Let the function $f(z) = z + a_3z^3 + a_5z^5 + \cdots$ be regular, univalent and odd for $|z| < 1$. Then all the sections $z + a_3z^3 + \cdots + a_{2n+1}z^{2n+1}$ $(n = 1, 2, \cdots)$ are univalent in the circle $|z| < 3^{-\frac{1}{2}}$ and the number $3^{-\frac{1}{2}}$ cannot be replaced by any larger one.

Joh proved this result in the case where all coefficients are real and in general for all values of n but $n = 3$. The values $n = 1, 2$ are treated as special cases, the former being trivial, the latter requiring the use of Löwner's expressions for the coefficients. We now give a general proof for the cases $n \geqq 3$. For this purpose we must recall the bounds $|a_9| < 1.38$, $|a_{11}| < 1.70$, $|a_{2n+1}| < 17/5$; cf. [7].

From the inequality (1) and the known bounds for the modulus of an odd univalent function we derive

$$|f(z_1) - f(z_2)|/|z_1 - z_2| > (1 - r^2)r^{-2}r^2(1 + r^2)^{-2} = 3/8;$$
$$|z_1| = |z_2| = r = 3^{-\frac{1}{2}}.$$

On the other hand,

$$\sum_{\nu=4}^{\infty} |a_{2\nu+1}|\,|z_1^{2\nu+1} - z_2^{2\nu+1}|/|z_1 - z_2|$$
$$< (1.38)(1/9) + (1.70)(11/243) + (17/5)\sum_{\nu=6}^{\infty}(2\nu + 1)3^{-\nu}.$$

This last number is less than .331 which in turn is less than 3/8 (= .375).

Now, as in § **2**, we see that all the sections for $n \geqq 3$ are univalent in $|z| < 3^{-\frac{1}{2}}$. That this bound is best possible is shown in [5].

The present method can also be used to simplify the proof of Joh's results on functions with bounded coefficients [6].

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] G. M. Golusin, "On the theory of univalent functions," *Matematicheskii Sbornik*, vol. 60 (1946), pp. 167-178. (Russian, English summary.)

[2] ———, "On distortion theorems and coefficients of univalent functions," *Matematicheskii Sbornik*, vol. 61 (1946), pp. 183-202. (Russian, English summary.)

[3] ———, "On the coefficients of univalent functions," *Matematicheskii Sbornik*, vol. 64 (1948), pp. 373-380 (Russian).

[4] H. Grötzsch, "Über die Verzerrung bei schlichter konformer Abbildung mehrfach zusammenhängender schlichter Bereiche," *Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften zu Leipzig*, Math-Phys Kl., vol. 81 (1929), pp. 38-47.

[5] K. Joh, "Theorems on the 'schlicht' functions," *Proceedings of the Physico-Mathematical Society of Japan*, vol. 19 (1937), pp. 1-12.

[6] ———, "Theorems on "schlicht" functions III," *ibid.*, vol. 21 (1939), pp. 191-208.

[7] V. Levin, "Some remarks on the coefficients of schlicht functions," *Proceedings of the London Mathematical Society*, vol. 39 (1935), pp. 467-480.

[8] ———, "Eine Bemerkung zum Koeffizientenproblem der schlichten Funktionen," *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 42 (1933), pp. 68-70.

[9] M. I. S. Robertson, "On the theory of univalent functions," *Annals of Mathematics*, vol. 37 (1936), pp. 374-408.

[10] G. Szegö, "Zur Theorie der schlichten Abbildungen," *Mathematische Annalen*, vol. 100 (1928), pp. 188-211.

ON NETS OF CURVES ON NON-ORIENTABLE SURFACES.*

By S. K. Zaremba.

1. Introduction.

The purely topological investigation of general nets of curves on closed surfaces was initiated by H. Hamburger [1] and B. v. Kerékjártó [2]. A further contribution by myself [4] was specifically concerned with orientable surfaces. However, nets of curves on non-orientable surfaces may also deserve attention, and their study is the object of the present paper.

We shall retain my former [4] definition of a *regular net of curves,* which is here re-stated for the reader's convenience:

1.1. Definition. S being any closed surface (*i. e.*, any closed two-dimensional variety), a set $\mathfrak{R}$ of curves on S will be called a *regular net of* curves on S, if to any point M on S (apart from a finite number of points, called *singular* with respect to $\mathfrak{R}$) there corresponds a neighborhood $U(M)$ of this point—called a *regular neighborhood*—capable of being topologically mapped on a neighborhood V of the Euclidean plane, so that the set of all the arcs of curves of $\mathfrak{R}$ contained (with the possible exception of their ends) in $U(M)$ be transformed in the set of all rectilinear segments parallel to either of two distinct fixed directions of the aforesaid plane and contained in V with the possible exception of their ends.

In what follows, $\mathfrak{R}$ will always denote a regular net of curves on a closed surface S. Obviously, it will not impair the generality of our reasoning if we limit ourselves to regular neighborhoods $U(M)$ bounded by closed curves.

1.2. Definition. According to the preceding definition, the arcs of curves of $\mathfrak{R}$ contained, possibly with the exception of their ends, in any regular neighborhood $U(M)$ can be so divided into two families that segments in the corresponding plane neighborhood V representing arcs of the same family be parallel to each other; such a partition will be called *normal.* If the whole net $\mathfrak{R}$ can be so decomposed into two families of curves that the resulting partition of all the arcs of curves of $\mathfrak{R}$ contained, with the possible exception of their ends, in any regular neighborhood be normal, $\mathfrak{R}$ will be called *monodrome* and the aforesaid non-local decomposition of $\mathfrak{R}$ into two families of curves will again be called *normal.*

---

* Received July 24, 1950.

This definition too repeats essentially the corresponding definitions contained in [3] and [4]. The same applies to the following definition, which, however, is now so formulated as to apply to non-monodrome as well as to monodrome nets of curves, which is necessary in the case of non-orientable surfaces.

1. 3. Definition. If all the curves of a regular net $\mathfrak{R}$ can be so oriented that in the plane neighborhood V corresponding to any regular neighborhood $U(M)$ the resulting orientation of the segments parallel to one direction be coherent, $\mathfrak{R}$ will be called *bilateral.* Similarly, in the case of a monodrome net, if the curves of one of its constituent families (corresponding to a normal partition of $\mathfrak{R}$) can be so oriented that in the plane neighborhood corresponding to any regular neighborhood $U(M)$ the resulting orientation of the rectilinear images of the arcs of curves of this family be coherent, the aforesaid family of curves will be called *bilateral.*

Whereas on an orientable surface, in the case of a monodrome net of curves, either both constituent families are bilateral or both not, on the contrary, on a non-orientable surface two families of curves obtained by a normal partition of a monodrome net cannot both be bilateral, but there are examples of monodrome nets of curves in which one of the two constituent families is bilateral. On the other hand, we shall see that (again in contrast with orientable surfaces, on which only monodrome nets of curves can be bilateral), on non-orientable surfaces non-monodrome nets are the only ones capable of being bilateral.

The contributions of H. Hamburger [1] and B. v. Kerékjártó [2] to the theory of nets of curves on surfaces were mainly based on a concept of index related to that of Kronecker and defined for a special kind of curvilinear polygons, each of which was supposed to be contained in one neighborhood. The idea of a non-local index, as introduced in [4], depends intrinsically on the orientation of the surface, so that it cannot be applied to the case we are now investigating. However, on non-orientable surfaces, it is possible to introduce in a similar way an index mod $\frac{1}{2}$ and, in the case of monodrome sets, even an index mod 1, which is sufficient to obtain, in particular, criteria of monodromy and bilaterality analogous to those previously found for nets of curves on orientable surfaces.

2. The index mod $\frac{1}{2}$. We shall need the following definition:

2. 1. Definition. By a *normal polygon* of the regular net of curves $\mathfrak{R}$ we mean any closed curvilinear polygon without multiple points, whose sides are arcs of curves of $\mathfrak{R}$, two consecutive sides being always contained in a common regular neighborhood $U(M)$ of their common vertex M.

Having chosen $U(M)$ and having carried out a normal partition of the arcs of curves of $\mathcal{R}$ contained, with the possible exception of their ends, in $U(M)$, we can determine whether the two sides of the given polygon meeting at M belong to the same family of curves or not (evidently, the result does not depend on the particular choice of $U(M)$); in the first case, M will be called a *neutral* vertex of the polygon.

In the second case, supposing the neighborhood $U(M)$, as well as the polygon itself, to be oriented, the vertex M can be labeled as positive or negative in the following way: We extend the two sides of the polygon adjacent to M as far as the boundary of $U(M)$ and so as to be still in the presence of arcs of curves of $\mathcal{R}$. Let LM and MN be these extended sides, their orientations from L to M and from M to N being consistent with that of the given polygon. Consider now that arc LN of the boundary of $U(M)$ which does not meet the extensions of LM and NM beyond M. If the orientation of this arc from L to N agrees with that induced by the orientation of $U(M)$, M will be reckoned as a *positive vertex*; otherwise M will be called *negative.*

In the case of an orientable surface, its orientation could be given, allowing to deduce therefrom the orientation of every $U(M)$, and the index of an oriented normal polygon a was defined by the formula

$$\text{ind}\,(a) = 1 - \tfrac{1}{4}s + \tfrac{1}{4}r, \tag{1}$$

where s was the number of its positive vertices and r that of its negative vertices [4]. The index of a normal polygon bounding a neighborhood of a singular point coincided with the index (according to Kerékjártó [2]) of this singular point whenever the orientation of the polygon was that induced by the orientation of the neighborhood.

If the surface is not orientable, this definition of the non-local index (as opposed to that of the index of a singular point, which is of local character) cannot be applied directly. We can, however, adopt an arbitrary orientation of a neighborhood of a starting point on a given normal polygon and propagate this orientation while describing the polygon in a given direction, thus giving a meaning to the index defined by (1). Obviously, if we return to the starting point with the opposite orientation, *i. e.*, if the cut on the surface corresponding to the given polygon is one-sided, the value found for the index depends on the starting point. More exactly, it is easy to see that, if the point moves so as to pass through a non-neutral vertex of the polygon, this point changes from positive to negative or conversely, so that the change in the value obtained for the index amounts to $\pm\,\frac{1}{2}$. Thus in the general case the index, whose value may be any multiple of $\frac{1}{4}$, is determined only mod $\frac{1}{2}$. It can therefore be defined in a simpler way as being equal, mod $\frac{1}{2}$, to one-quarter of the number of non-neutral vertices of the given polygon. This amounts to a definition of the values of the index as residue classes modulo one-half. Evidently, the value of the index of a normal polygon, as defined above, does not depend on its orientation; this is why, in what follows, topological complexes will be considered only mod 2. If a is any normal polygon, its index modulo one-half will be denoted by $\mathrm{ind}_{\frac{1}{2}}(a)$.

It follows from our definition that if the normal polygon a bounds a neighborhood containing exactly one singular point, let us say P, the residue class mod $\frac{1}{2}$ of the index of P as defined by Kerékjártó [2] is $\mathrm{ind}_{\frac{1}{2}}(a)$. In particular, the index mod $\frac{1}{2}$ of a normal polygon bounding a neighborhood without singular points is zero.

2. 2. THEOREM. *Let a and b be two normal polygons whose common part forms an open polygon (which can always be assumed to be bounded by two vertices common to a and b), so that their sum* mod 2 *(in the sense of the theory of topological complexes), which will be denoted by $a + b$, be another normal polygon. Then* $\mathrm{ind}_{\frac{1}{2}}(a) + \mathrm{ind}_{\frac{1}{2}}(b) = \mathrm{ind}_{\frac{1}{2}}(a + b)$.

Proof. The contribution of any vertex belonging to one only of the polygons a and b is the same for both sides of this equality. The contributions of any interior vertex of the common part of a and b to $\mathrm{ind}_{\frac{1}{2}}(a)$ and to $\mathrm{ind}_{\frac{1}{2}}(b)$ are the same, and since they amount to 0 or $\frac{1}{4}$, they cancel each other modulo $\frac{1}{2}$ in $\mathrm{ind}_{\frac{1}{2}}(a) + \mathrm{ind}_{\frac{1}{2}}(b)$, whereas in $a + b$ the same vertex does not appear at all. Finally, any of the two ends of the common part of a and b must obviously be a non-neutral vertex either in both polygons a and b or in one of them only; in the former case, and only in that case, such a vertex is neutral in $a + b$, so that in both cases its contributions to $\mathrm{ind}_{\frac{1}{2}}(a) + \mathrm{ind}_{\frac{1}{2}}(b)$ and to $\mathrm{ind}_{\frac{1}{2}}(a + b)$ are congruent mod $\frac{1}{2}$. Thus adding together mod $\frac{1}{2}$ all these contributions to $\mathrm{ind}_{\frac{1}{2}}(a) + \mathrm{ind}_{\frac{1}{2}}(b)$ on the one hand and to $\mathrm{ind}_{\frac{1}{2}}(a + b)$ on the other, we find the same residue classes. Hence the proof is complete.

By modifying slightly the argument in [4], § 2. 5, to suit topological complexes mod 2, we obtain the two following lemmas:

2. 3. LEMMA. *If k is any component of the common part of two normal polygons, c_1 and c_2, and is bounded by two distinct vertices common to c_1 and c_2, the sum $c_1 + c_2$, treated as a complex* mod 2, *can be resolved into a finite number of normal polygons, say $a_1, \cdots, a_n$, composed of sides of c_1 and c_2 with the exclusion of those belonging to k, so that no side occurs in more than two of the polygons $a_1, \cdots, a_n$.*

2. 4. LEMMA. *On the assumption of the preceding lemma,*

$$\operatorname{ind}_{\frac{1}{2}}(c_1) + \operatorname{ind}_{\frac{1}{2}}(c_2) = \sum_{i=1}^{i=n} \operatorname{ind}_{\frac{1}{2}}(a_i).$$

Hence, by mathematical induction with respect to the total number of curvilinear segments involved, we deduce the following theorem.

2. 5. THEOREM. *If $c_1, \cdots, c_m$ are normal polygons whose sum* mod 2 *is the zero complex,*

$$\sum_{i=1}^{i=n} \operatorname{ind}_{\frac{1}{2}}(c_i) = 0.$$

2. 6. COROLLARY. *If $a_1, \cdots, a_m$ and $b_1, \cdots, b_n$ are normal polygons such that the complexes* mod 2 *denoted by $a_1 + \cdots + a_m$ and $b_1 + \cdots + b_n$ respectively are equal,*

$$\sum_{i=1}^{i=m} \operatorname{ind}_{\frac{1}{2}}(a_i) = \sum_{i=1}^{i=n} \operatorname{ind}_{\frac{1}{2}}(b_i).$$

Finally, the following theorem, corresponding to 2 · 9 in [4], can be stated and proved in a simplified way owing to the replacement of ordinary integers by residue classes:

2. 7. THEOREM. *If s is the remainder class* mod $\frac{1}{2}$ *of the sum of the indices of the singular points contained in a part of the surface S bounded by the normal polygons $a_1, \cdots, a_m$,*

$$s = \sum_{i=1}^{i=m} \operatorname{ind}_{\frac{1}{2}}(a_i).$$

Proof. This part of S can be decomposed into a finite number of neighborhoods, none of which contains more than one singular point, and which are bounded by some normal polygons, say $b_1, \cdots, b_n$. Clearly

$$s = \sum_{i=1}^{i=n} \operatorname{ind}_{\frac{1}{2}}(b_i).$$

On the other hand, $a_1 + \cdots + a_m = b_1 + \cdots + b_n$ and therefore, by 2. 6,

$$s = \sum_{i=1}^{i=n} \operatorname{ind}_{\frac{1}{2}}(b_i) = \sum_{i=1}^{i=m} \operatorname{ind}_{\frac{1}{2}}(a_i).$$

Now criteria for the monodromy (see 1. 2) of regular nets of curves can readily be obtained.

2. 8. THEOREM. *$\mathfrak{R}$ is monodrome, if, and only if, the indices* mod $\frac{1}{2}$ *of all the normal polygons on S are* 0.

This theorem is the counterpart of the result of 2. 31 in [3] and of 3. 1 in [4].

Proof. Wishing to divide $\mathfrak{R}$ into two families, say $\mathfrak{A}$ and $\mathfrak{B}$, let us start from any regular neighborhood $U(M)$, where the arcs of the curves of $\mathfrak{R}$ have undergone a normal partition into two families, $\mathfrak{A}$ and $\mathfrak{B}$, and let us propagate this partition similarly to the analytic continuation of functions of complex variables. Obviously, $\mathfrak{R}$ is monodrome if, and only if, we always come back to $U(M)$ with the initial determination of $\mathfrak{A}$ and $\mathfrak{B}$, and the path followed can evidently be supposed to be a normal polygon. Now, when tracing a normal polygon, we meet non-neutral vertices when, and only when, we pass from an arc of a curve $\mathfrak{A}$ to an arc of a curve of $\mathfrak{B}$ or conversely, so that we return with the initial partition of $\mathfrak{R}$ if and only if the number of non-neutral vertices is even, *i. e.*, if and only if the index mod $\frac{1}{2}$ is 0.

As it may be preferable not to have to consider all possible normal polygons, let us finally replace the preceding proposition by a more special criterion, directly deducible from 2. 7 and 2. 8.

2. 9. THEOREM. *A regular net of curves $\mathfrak{R}$ on a surface S is monodrome if and only if the indices* mod $\frac{1}{2}$ *of the normal polygons of a set forming a base of homology for S, as well as the residues* mod $\frac{1}{2}$ *of the indices of the singular points are all* 0.

3. The index mod 1. Throughout this section, we assume the monodromy of $\mathfrak{R}$, and denote by $\mathfrak{A}$ and $\mathfrak{B}$ the two families of curves obtained as a result of a normal decomposition of $\mathfrak{R}$.

3. 1. DEFINITION. On this assumption, if a normal polygon a is given, and if we apply the procedure described in 2. 1, starting from any interior point of a side a belonging to $\mathfrak{A}$, the result thus obtained will be determined mod 1. Indeed, if the starting point remains on the same side of a, nothing will be changed in the computation of the index. If, however, this point moves to the next side of a belonging to $\mathfrak{A}$, it has to pass through two non-neutral vertices. The contribution of each of them to the index changes by $\pm \frac{1}{2}$ if the cut determined by a is one-sided, and does not change at all if this cut is two-sided; so that the value obtained for the index is either unchanged or changed by ± 1, leaving the corresponding residue class mod 1 unchanged in any case. Hence, if the starting point moves from one side of a belonging to $\mathfrak{A}$ to any other side also belonging to $\mathfrak{A}$, this residue class will remain the same. It will be called the *index* mod 1 *of a with respect to $\mathfrak{A}$* and will be denoted by $\mathrm{ind}_{\mathfrak{A}}(a)$.

According to 2. 8, the only possible values for $\mathrm{ind}_{\mathfrak{A}}(a)$ are the residue classes of 0 and $\frac{1}{2}$, so that $-\mathrm{ind}_{\mathfrak{A}}(a) = \mathrm{ind}_{\mathfrak{A}}(a)$. Hence the index mod 1 depends neither on the orientation of a nor on that of the initial neighborhood.

3. 2. *Remark.* On the other hand, if the cut determined by a is one-sided, $\operatorname{ind}_{\mathfrak{B}}(a) = \operatorname{ind}_{\mathfrak{A}}(a) + \frac{1}{2}$. This property of the index mod 1 is in contrast with the corresponding properties both of the index mod $\frac{1}{2}$ and of the index defined in [4] for orientable surfaces.

3. 3. THEOREM. *If a and b are two normal polygons whose common part is an open polygon bounded by two vertices common to a and b, so that $a + b$ is another normal polygon,*

$$\operatorname{ind}_{\mathfrak{A}}(a) + \operatorname{ind}_{\mathfrak{A}}(b) = \operatorname{ind}_{\mathfrak{A}}(a + b).$$

Proof. The polygons a and b can be supposed to be so oriented that the resulting orientations of their common part be opposite. The orientations of the regular neighborhoods covering together both polygons can so be determined that they coincide on the common part of a and b. On these assumptions, it is easy to see that each of the two ends of the common part of a and b either is non-neutral in both a and b and has the same sign in a as in b, but becomes neutral in $a + b$, or is non-neutral in one only of the polygons a and b and remains so in $a + b$, but with the opposite sign. Thus the contribution of each of these two vertices to $\operatorname{ind}_{\mathfrak{A}}(a + b)$ differs from that to $\operatorname{ind}_{\mathfrak{A}}(a) + \operatorname{ind}_{\mathfrak{A}}(b)$ by $\pm \frac{1}{2}$ and both these differences, added together mod 1, cancel each other. On the other hand, the contributions to $\operatorname{ind}_{\mathfrak{A}}(a) + \operatorname{ind}_{\mathfrak{A}}(b)$ of any other vertex common to a and b cancel each other, while the vertex itself disappears in $a + b$. Finally, each of the vertices occurring in one only of the polygons a and b yields the same contribution to both sides of the equality to be proved. Hence $\operatorname{ind}_{\mathfrak{A}}(a) + \operatorname{ind}_{\mathfrak{A}}(b) = \operatorname{ind}_{\mathfrak{A}}(a + b)$.

The preceding theorem can be generalized as follows:

3. 4. LEMMA. *On the assumptions of Lemma 2. 3, if $\mathfrak{R}$ is monodrome and split up into two families of curves, $\mathfrak{A}$ and $\mathfrak{B}$ (this partition being normal),*

$$\operatorname{ind}_{\mathfrak{A}}(c_1) + \operatorname{ind}_{\mathfrak{A}}(c_2) = \sum_{i=1}^{i=n} \operatorname{ind}_{\mathfrak{A}}(a_i).$$

Hence, by mathematical induction with respect to the total number of curvilinear segments involved, we deduce the following theorem:

3. 5. THEOREM. *If $c_1, \cdots, c_m$ are normal polygons whose sum* mod 2 *is the zero complex,*

$$\sum_{i=1}^{i=m} \operatorname{ind}_{\mathfrak{A}}(c_i) = 0.$$

3. 6. COROLLARY. *If $a_1, \cdots, a_m$ and $b_1, \cdots, b_n$ are normal polygons such that the complexes* mod 2 *represented by $a_1 + \cdots + a_m$ and $b_1 + \cdots + b_n$ are equal,*

$$\sum_{i=1}^{i=m} \operatorname{ind}_{\mathfrak{a}}(a_i) = \sum_{i=1}^{i=n} \operatorname{ind}_{\mathfrak{a}}(b_i).$$

Corresponding to 2. 7, the following can be proved.

3. 7. THEOREM. *If s is the residue class* mod 1 *of the sum of the indices of the singular points contained in a part of S bounded by the normal polygons $a_1, \cdots, a_m$,*

$$s = \sum_{i=1}^{i=m} \operatorname{ind}_{\mathfrak{a}}(a_i).$$

The following criteria for the bilaterality (see 1. 3) of a family of curves are analogous to those for the monodromy of a net of curves (see § 2).

3. 8. THEOREM. *The family $\mathfrak{a}$ is bilateral if and only if the indices* mod 1 *with respect to $\mathfrak{a}$ of all the normal polygons on S are* 0.

Proof. The proof is similar to that in [4], § 3. 5.

In order to give an orientation to the curves of $\mathfrak{a}$, let us start from any normal neighborhood $U(M)$, where the arcs of the curves of $\mathfrak{a}$ are oriented in a coherent way and propagate this partition so that, in the part common to two overlapping neighborhoods, the corresponding orientations of the curves of $\mathfrak{a}$ agree with each other.

$\mathfrak{a}$ is bilateral if and only if we always return to the initial neighborhood with the same orientation of the curves of $\mathfrak{a}$, whenever we describe a normal polygon. In agreement with 3. 1, the starting point can always be supposed to be on a side of the polygon formed by an arc of a curve of $\mathfrak{a}$. Such a side will be called an $\mathfrak{a}$-side and will be regarded as *positive* if its orientation derived from that of the family agrees with that of the polygon; otherwise it will be regarded as *negative*. If two consecutive $\mathfrak{a}$-sides have the same sign, they are separated either by a neutral vertex or by two vertices of opposite signs, possibly together with an unspecified number of neutral vertices; in both cases, the total contribution of these vertices to the index of the polygon is 0. On the other hand, it is easy to see that two consecutive $\mathfrak{a}$-sides of opposite signs are separated by two vertices of the same sign, possibly together with an unspecified number of neutral vertices, giving together a contribution to the index of the polygon amounting to $\pm \frac{1}{2}$.

After having traced the polygon, we come back to the starting point with the original orientation of the curves of $\mathfrak{a}$ if and only if we have passed as many times from a positive $\mathfrak{a}$-side to a negative side as from a negative side to a positive one; that is to say (according to what we have just seen), if and only if the value found for the index following the definition 3. 1 is an integer. Hence the proof is complete.

As a direct consequence of the preceding theorems, we find the following criterion of bilaterality:

3. 9. THEOREM. *The family $\mathcal{A}$ is bilateral if and only if the indices* mod 1 *with respect to $\mathcal{A}$ of a set of normal polygons forming a base of homology for S, as well as the residues* mod 1 *of the indices of the singular points are all* 0.

4. Bilateral nets of curves. This paper deals mainly with nets of curves on non-orientable surfaces, but the propositions established above are valid for orientable surfaces as well. In order to show the contrast between the cases of orientable and non-orientable surfaces, we shall first deduce from these propositions the following theorem:

4. 1. THEOREM. *A bilateral net of curves on an orientable surface is necessarily monodrome.*

Proof. Let a be a normal polygon with respect to a bilateral net, $\mathcal{R}$, of curves on an orientable surface S. We denote by M any point on a other than a vertex and we choose an orientation of S arbitrarily. Starting from M, let us follow a in any of the two directions, counting on our way, according to formula (1), the contributions of the various vertices to the index of a and propagating, at the same time, from neighborhood to neighborhood, a normal partition of $\mathcal{R}$ into two families of curves, say $\mathcal{A}$ and $\mathcal{B}$, as well as an orientation of $\mathcal{R}$, both initially determined in a regular neighborhood $U(M)$ of M. Returning to M for the first time, we may bring with us the initial determination of $\mathcal{A}$ and $\mathcal{B}$, or else the two families can be permuted. However, if we continue following a in the same direction and propagating both the orientation of $\mathcal{R}$ and its partition into $\mathcal{A}$ and $\mathcal{B}$, the next time we reach M again, the determination of $\mathcal{A}$ and $\mathcal{B}$ will be the initial one. $\mathcal{R}$ being bilateral, $\mathcal{A}$ and $\mathcal{B}$ will also be oriented as before. Hence, by an argument similar to that of 3. 8, we see that if we keep counting on our way the contributions of the encountered vertices to the index of a, the total with which we reach M after two circuits must be an integer. Since, owing to the orientability of S, each vertex appears both times with the same sign, this total is twice the sum obtained after one circuit. This sum is therefore a multiple of $\frac{1}{2}$, which means that $\mathrm{ind}_{\frac{1}{2}}(a) = 0$. The normal polygon a being arbitrary, this implies, according to 2. 8, that $\mathcal{R}$ is monodrome.

4. 2. It is also clear that if a monodrome net of curves on an orientable surface is normally decomposed into two families, both these families are simultaneously bilateral or not.

The converse of 4. 1 and 4. 2 is true in the case of non-orientable surfaces.

Indeed, since on a non-orientable surface covered by a regular net of curves there are always normal polygons determining one-sided cuts, 3. 2 and 3. 8 imply the following proposition.

4. 3. PROPOSITION. *If a monodrome net of curves on a non-orientable surface is subject to a normal decomposition into two families of curves, both of them cannot be bilateral.*

This can also be deduced from the fact that a coherent orientation of both these families of curves would define a coherent orientation of the surface.

4. 4. However, one of the two families of curves obtained by the normal decomposition of a net on a non-orientable surface can be bilateral. A simple example of such a net on the projective plane is afforded by a pencil of concentric circles supplemented by the straight line at infinity, the other family consisting of the diameters common to all the circles of the pencil. The possibility of imparting a coherent orientation to the first family is obvious, and it is not difficult to see that the family composed of the diameters is not bilateral.

4. 5. On the other hand, non-monodrome nets of curves on non-orientable surfaces can be bilateral. This is shown by the following example:

We obtain a Klein bottle from the square Q of the (x, y)-plane defined by the inequalities $0 \leq x \leq 2\pi$, $0 \leq y \leq 2\pi$, if we identify the points $(x, 0)$ and $(x, 2\pi)$ for any value of x, and the points $(0, y)$ and $(2\pi, 2\pi - y)$ for any value of y. On this surface, the characteristics of the differential equation

$$\sin \tfrac{1}{2}x \, dx^2 + 2 \cos \tfrac{1}{2}x \, dxdy - \sin \tfrac{1}{2}x \, dy^2 = 0$$

or

$$(2) \qquad (\cos \tfrac{1}{4}x \, dx - \sin \tfrac{1}{4}x \, dy)(\sin \tfrac{1}{4}x \, dx + \cos \tfrac{1}{4}x \, dy) = 0$$

form a regular net of curves without singular points.

The fact that this net is bilateral (and therefore also not monodrome) can best be seen as follows: The two orthogonal vectors u and v of coordinates $(\sin \frac{1}{4}x, \cos \frac{1}{4}x)$ and $(\cos \frac{1}{4}x, -\sin \frac{1}{4}x)$ respectively are tangent to the characteristics of (2) at any point of Q and locally they determine an orientation of the curves of the net. Since they are continuous, in order to prove that this orientation is coherent, we have only to verify that the orientations of the curves determined by these vectors agree on both pairs of sides of Q which are identified. But this is clear as far as the sides $y = 0$ and $y = 2\pi$ are concerned, the identified points having the same x-value, which alone determines the vectors. On the side $x = 0$, the vectors u and v become $(0, 1)$ and $(1, 0)$, whereas on the side $x = 2\pi$ they become $(1, 0)$ and $(0, -1)$, respectively. Owing to the manner of identifying these two sides, the vector $(0, 1)$ on $x = 0$ coincides with $(0, -1)$ on $x = 2\pi$, so that the two vectors u and v are only subjected to a permutation when we cross the meridian corresponding to the two sides $x = 0$ and $x = 2\pi$ of Q. This permutation shows that the net is not monodrome, and the fact that otherwise the directions of the pairs of vectors agree with each other proves that the net is bilateral, one of its two possible orientations being determined by the couple of vectors u and v.

It should be noted that in general the definition of the index by means of formula (1), such as it was stated in [4] and recalled above (cf. 2. 1), can be applied, in the case of a non-orientable surface, not only to normal polygons bounding neighborhoods (for instance, those of singular points), but also to any normal polygons corresponding to two-sided cuts on the given surface. The value of the index then depends on the orientation of the polygon and on that of the surface in its neighborhood, but the question of its being an integer or not is clearly independent of both these orientations (see also [4], § 2. 2). This leads to conclude with the following criterion of bilaterality for regular nets of curves, which applies to orientable and non-orientable surfaces alike and includes, as a particular case, a previously stated criterion of bilaterality for nets of curves on orientable surfaces (cf. [4], § 3. 5).

4. 6. THEOREM. *A regular net of curves on a surface S is bilateral if and only if the indices of normal polygons corresponding to two-sided cuts of S are all integers, whereas the indices* mod $\frac{1}{2}$ *of normal polgons corresponding to one-sided cuts of S are all equal to* $\frac{1}{4}$.

Proof. (i) The condition is necessary. With respect to polygons corresponding to two-sided cuts of S, the arguments of 4. 1 and 3. 8 can be applied to show that the indices must all be integers. On the other hand, if a normal polygon corresponding to a one-sided cut of S had an index mod $\frac{1}{2}$ equal to 0, then, by the argument of 2. 8, the net of curves could be split up normally into two families of curves in a neighborhood of the polygon; so that the definition 3. 1 of the index mod 1 could be applied to each of these two families. But according to 3. 2, the indices mod 1 of the polygon with respect to the two families cannot vanish simultaneously. Consequently, at least one of these two families must fail to be bilateral, which precludes the bilaterality of the whole net.

(ii) The condition is sufficient. In an arbitrarily chosen regular neighborhood $U(M)$, let us split up normally the given net of curves into two families and determine the orientation of one of these two families, which will be denoted by $\mathfrak{A}$. The determination of the family $\mathfrak{A}$, as well as its orientation, can be propagated from neighborhood to neighborhood through their overlapping parts. The case of an orientable surface having been dealt with previously, we can suppose S to be non-orientable, so that there exist normal polygons corresponding to one-sided cuts. According to the argument of 2. 8, if, following such a polygon, we propagate the determination of $\mathfrak{A}$ together with its orientation, we shall return to $U(M)$ with a changed determination of the family $\mathfrak{A}$. This will precisely give the orientation of the second family of curves in $U(M)$, thus enabling us to orient the whole net of curves, provided that we are not led to inconsistencies.

The consistency of the aforesaid way of determining the orientation of the given net of curves will be proved if we show that any closed path on the surface has either of the two following properties:

(α) if we follow it, starting from any neighborhood where the family $\mathfrak{A}$, as well as its orientation, is locally determined, we come back with the initial determination and the initial orientation of $\mathfrak{A}$;

(β) if we trace it once, starting from any point of a neighborhood where both the family $\mathfrak{A}$ and its orientation are locally given, we return with a different determination of $\mathfrak{A}$, but if we trace the same path once more in the same direction, eventually we arrive at the starting point with the initial determination and the initial orientation of $\mathfrak{A}$.

Obviously, we can always suppose the path to be formed by a regular polygon of the net of curves. If it corresponds to a two-sided cut of the surface, its index is an integer, and the argument of 3. 8 shows that the polygon has the property (α).

If, on the other hand, the polygon corresponds to a one-sided cut of the surface, its index mod $\frac{1}{2}$ is $\frac{1}{4}$. Consequently, as is shown by the proof of 2. 8, after one tour, we reach the initial point with a changed determination of $\mathfrak{A}$, *i. e.*, the two families into which the net is locally split up are permuted. Therefore, after having traced the polygon once more in the same direction, we recover the initial determination of $\mathfrak{A}$, and we have only to prove that we return with the initial orientation of $\mathfrak{A}$.

We can always suppose the starting point to be on a side formed by an arc of a curve of $\mathfrak{A}$. If an orientation of the initial neighborhood has been chosen, it can be propagated along the polygon with the determination and the orientation of $\mathfrak{A}$; so that, while following the polygon, we can count the contributions of its vertices to the index, in accordance with formula (1). It is clearly possible to discard neutral vertices on the polygon, that is to say, to consider two sides connected by a neutral vertex as one side. Then, while we follow the polygon, every other side is labeled as belonging to $\mathfrak{a}$ and, as such, provided with an orientation. This orientation may coincide or not with that of the polygon. As in the proof of 3. 8, such a side will be called *positive* in the former case and *negative* in the latter. Between two consecutive sides considered as belonging to $\mathfrak{a}$ there are two vertices and, as in 3. 8, it is easy to see that these vertices are of the same sign if and only if the aforesaid sides are of opposite signs. Hence the joint contribution to the index of two vertices situated between two sides attributed to $\mathfrak{a}$ is 0 if these sides have the same sign and equals $\pm \frac{1}{2}$ if they are of opposite signs.

On the other hand, if we trace the polygon twice, we pass through each vertex twice, but (owing to the one-sidedness of the corresponding cut) with opposite orientations of its neighborhood, so that the same vertex appears once as positive and once as negative, and its total contribution to the index is 0. The total value for the index is therefore an integer (it is immaterial whether or not we count the constant term in (1)). This shows that we pass as many times from a positive side to a negative side as from a negative side to a positive one. In other terms, after describing the polygon twice in the same direction, we return with the initial orientation of the family $\mathfrak{a}$, i. e., the polygon has the property (β). Hence the proof is complete.

POLISH UNIVERSITY COLLEGE, LONDON, ENGLAND.

REFERENCES.

[1] H. Hamburger, "Über Kurvennetze mit isolierten Singularitäten auf geschlossenen Flächen, Teil I," *Mathematische Zeitschrift*, vol. 19 (1924), pp. 50-66. The second part of this paper has not been published.

[2] B. v. Kerékjártó, *Topologie*, I (Berlin, 1923), pp. 250-261.

[3] S. K. Zaremba, "Théorie des réseaux quasi-réguliers, I," *Ann. Soc. Polon, Math.*, vol. 14 (1936), pp. 1-73.

[4] ———, "Contribution à théorie des réseaux sur les surfaces fermées," *Bull. de l'Acad. Polon. des Sci.*, ser. A, 1939, pp. 133-149 (published in 1946).

GENERALIZED STIELTJES-POST INVERSION FORMULA FOR INTEGRAL TRANSFORMS INVOLVING A PARAMETER.*

By L. C. Hsu.

---

1. **Introduction.** The object of this paper is to establish an inversion formula for a kind of integral transforms involving a parameter. The classical Stieltjes-Post inversion formula for Laplace transform is included as a particular case. Our method is an extension of the method of Laplace, Haviland [3] and Widder.

Let $f(x)$ be the real Laplace transform of $g(u)$, namely,

$$(1) \qquad f(x) = \int_0^\infty e^{-xu} g(u)\, du,$$

where the right-hand side of (1) converges for some x and $g(u) \varepsilon L(o, R)$ for every positive R. It is a well-known result discovered by Stieltjes and Post and generalized by Widder (see [1; pp. 276-289], [2; pp. 723-781]) that, for every positive t in the Lebesgue set for $g(u)$, $g(t)$ can be expressed in terms of the derivative $f^{(\lambda)}(\lambda/t)$ with $\lambda \to +\infty$, viz.

$$(2) \qquad g(t) = \lim_{\lambda\to\infty} (-1)^\lambda (\lambda/t)^{\lambda+1} f^{(\lambda)}(\lambda/t)/\lambda!.$$

By Stirling's formula we may rewrite (2) as

$$(2)^1 \qquad g(t) = \lim_{\lambda\to\infty} (-1)^\lambda (2\lambda\pi)^{-\frac{1}{2}} (e/t)^\lambda (\lambda/t) f^{(\lambda)}(\lambda/t).$$

Since (1) can be differentiated any number of times (under the integral sign) within its region of convergence [1; p. 57], we have

$$(3) \qquad (-1)^\lambda f^{(\lambda)}(\lambda/t) = \int_0^\infty e^{-\lambda u/t} u^\lambda g(u)\, du, \qquad (\lambda = 1, 2, 3, \cdots).$$

Hence (2) or $(2)^1$ may also be considered as an inversion formula for the integral transforms defined by (3).

Instead of the special transforms of (3), we now consider integral transforms of the form

$$(4) \qquad f(\lambda, t) = \int_0^\infty \psi(u, \lambda, t) g(u)\, du,$$

where λ is a real parameter and the right-hand side of (4) is assumed to be convergent for some $t > 0$, and $\psi(u, \lambda, t)$ is a real valued function of

---

* Received May 23, 1950.

$u, \lambda, t\,(u \geqq 0, \lambda > 0, t > 0)$ having positive values when $u, \lambda, t > 0$. As in the ordinary case we call f, g, ψ the generating function, determining function and transform function respectively.

Under certain conditions we shall prove a general inversion formula of the form

$$(5)\quad g(\theta(t)) = \lim_{\lambda\to\infty}(\{-(\partial/\partial u)^2 \log\psi(u,\lambda,t)\}/\{2\pi(\psi(u,\lambda,t))^2\})^{\frac{1}{2}}_{u=\phi(\lambda,t)} f(\lambda,t),$$

where $\phi(\lambda, t)$ is a function satisfying the equation

$$(6)\qquad [\partial/\partial u \log\psi(u,\lambda,t)]_{u=\phi(\lambda,t)} = 0$$

for all large values of λ and certain $t > 0$, and

$$(7)\qquad \lim_{\lambda\to\infty}\phi(\lambda,t) = \theta(t) > 0.$$

The relation (7) may be used for determining $\theta(t)$.

It is easy to verify that the Stieltjes-Post inversion formula $(2)^1$ is implied by (5). For if we take $\psi(u,\lambda,t) = e^{-\lambda u/t}u^{\lambda}$, then $\partial/\partial u \log\psi(u,\lambda,t) = \lambda(1/u - 1/t)$, $(\partial/\partial u)^2 \log\psi(u,\lambda,t) = -\lambda u^{-2}$, so that $u = t$ is the only solution satisfying (6), and consequently $\phi(\lambda,t) = \theta(t) = t$. Hence a direct substitution into (5) yields the results $(2)^1$.

2. Some lemmas. In order to investigate sufficient conditions under which the general formula (5) is valid, we need some lemmas concerning the asymptotic evaluation of integrals. In what follows we write $(\partial/\partial u)F(u,v)$ $F_u(u,v)$, $(\partial/\partial u)^2F = F_{uu}$, etc. A will be used to denote a certain positive number not necessarily the same in different occurrences.

LEMMA 1. *Let* $F(u,v)$, $F_u(u,v)$, $F_{uu}(u,v)$ *be continuous functions defined in the domain* D $(-\infty \leqq a \leqq u \leqq b \leqq \infty,\ v > 0)$ *such that*

(i) *there is a continuous function* $u = \phi(v)$ *satisfying* $F_u(\phi(v), v) \equiv 0$ *for all large* v, *and* $\phi(v) \to \xi$ $(a < \xi < b)$ *as* $v \to \infty$;

(ii) $F_{uu}(u,v) < 0$ *for all* u $(a < u < b)$ *and large* v;

(iii) $F_{uu}(u,v) \sim F_{uu}(\xi,v) \to -\infty$ *as* $u \to \xi$, $v \to \infty$ *simultaneously.*[1] *Then as* $\lambda \to \infty$ *we have*

$$(8)\qquad \int_a^b \exp[F(u,\lambda)]\,du \sim \exp[F(\phi(\lambda),\lambda)](-2\pi/F_{uu}(\xi,\lambda))^{\frac{1}{2}}.$$

---

[1] Precisely speaking, $F_{uu}(u,v) \sim F_{uu}(\xi,v)$ means that for any given $\epsilon > 0$ there exist a small number $\delta = \delta(\epsilon) > 0$ and a large number $M = M(\epsilon)$ such that $|F_{uu}(u,v)/F_{uu}(\xi,v) - 1| < \epsilon$ whenever $|u - \xi| < \delta$, $v > M$. Similarly $F_{uu}(u,v) \to -\infty$ means that $-F_{uu}(u,v) > N$ for any given N, provided that $|u - \xi| < \delta$, $v > M$, where δ, M depend on N.

Proof. It suffices to give a proof for the case $a = -\infty$, $b = +\infty$. By hypotheses (i) and (ii) it is obvious that for every large $v = \lambda$ (say $\lambda > N$) the integrand function $\exp[F(u, \lambda)]$ always assumes a maximum at $u = \phi(\lambda)$, since $F_u(\phi(\lambda), \lambda) = 0$ and $F_{uu} < 0$. Moreover, hypotheses (i) and (ii) ensure that the implict function theorem is applicable to $F_u(u, v) = 0$, so that $u = \phi(v)$ is actually a single-valued function of v.

Fix λ and set

$$\int_{-\infty}^{+\infty} \exp[F(u,\lambda)]du = \int_{-\infty}^{\phi(\lambda)-\epsilon(\lambda)} + \int_{\phi(\lambda)-\epsilon(\lambda)}^{\phi(\lambda)+\epsilon(\lambda)} + \int_{\phi(\lambda)+\epsilon(\lambda)}^{+\infty} = J_1 + J_2 + J_3,$$

where $\epsilon(\lambda) > 0$ is a certain function tending to zero with $1/\lambda$. By Taylor's theorem with remainder we have

$$J_2(\lambda)\exp[-F(\phi(\lambda),\lambda)] = \int_{\phi(\lambda)-\epsilon(\lambda)}^{\phi(\lambda)+\epsilon(\lambda)} \exp[\tfrac{1}{2}(u-\phi(\lambda))^2 F_{uu}(X,\lambda)]du,$$

where $X = X(\lambda)$ and $|X(\lambda) - \phi(\lambda)| < \epsilon(\lambda)$. Set

$$M(\lambda) = \underset{\phi(\lambda)-\epsilon<x<\phi(\lambda)+\epsilon}{\text{l. u. b.}} |F_{uu}(x,\lambda)|, \quad m(\lambda) = \underset{\phi(\lambda)-\epsilon<x<\phi(\lambda)+\epsilon}{\text{g. l. b.}} |F_{uu}(x,\lambda)|.$$

Then by a simple change of variable and upon a comparison with the probability integral we easily obtain

$$(9) \qquad (1-\delta)(2\pi/M(\lambda))^{\frac{1}{2}} \leq J_2(\lambda)\exp[-F(\phi(\lambda),\lambda)] \leq (2\pi/m(\lambda))^{\frac{1}{2}},$$

where

$$(1-\delta)(2\pi)^{\frac{1}{2}} = \int_{-\epsilon(M(\lambda))^{\frac{1}{2}}}^{\epsilon(M(\lambda))^{\frac{1}{2}}} e^{-\frac{1}{2}t^2}dt.$$

Thus $\delta = \delta(\lambda) \to 0$ provided that $\epsilon(\lambda)(M(\lambda))^{\frac{1}{2}} \to \infty$ with λ. Clearly hypothesis (iii) implies that

$$M(\lambda)/m(\lambda) \to 1, \qquad M(\lambda)/(-F_{uu}(\phi(\lambda),\lambda)) \to 1, \qquad (\lambda \to \infty),$$

since $\epsilon(\lambda) \to 0$ and $\phi(\lambda) \to \xi$. Thus in order to prove Lemma 1 we need only show that there exists a positive function $\epsilon(\lambda)$ such that when $\lambda \to \infty$, we have

I) $\epsilon(\lambda) \to 0$, II) $\epsilon(\lambda)(-F_{uu}(\phi(\lambda),\lambda))^{\frac{1}{2}} \to \infty$,

III) $J_1(\lambda)/J_2(\lambda) \to 0$, IV) $J_3(\lambda)/J_2(\lambda) \to 0$.

For the purpose just mentioned we may take

$$(10) \qquad \epsilon(\lambda) = (\log[-F_{uu}(\phi(\lambda),\lambda)]/(-F_{uu}(\phi(\lambda),\lambda)))^{\frac{1}{2}}.$$

That the conditions I) and II) are fulfilled is obvious by our hypothesis (iii). We now only verify the conidtion IV), as the justification of III) is similar.

Obviously for every large λ, $F(u,\lambda)$ is a monotonic decreasing function in $(\phi(\lambda),\infty)$, for, by hypothesis (ii) we have $F_{uu}(u,\lambda) < 0$ and consequently $F_u(u,\lambda) < 0$ in $(\phi(\lambda),\infty)$. Thus making change of variable $F(u,\lambda) = -t$ in J_3, setting $F(\phi(\lambda) + \epsilon(\lambda),\lambda) = T$ and using the law of the mean we obtain

$$\begin{aligned} 0 < J_3(\lambda) &\leq \int_{-T}^{+\infty} \{-F_u(\phi(\lambda) + \epsilon(\lambda),\lambda)\}^{-1} \exp(-t)\,dt \\ &= \int_{-T}^{+\infty} \{-\epsilon(\lambda) F_{uu}(\phi(\lambda) + \theta\epsilon(\lambda),\lambda)\}^{-1} \exp(-t)\,dt \\ &= \{-\epsilon(\lambda) F_{uu}(\phi(\lambda) + \theta\epsilon,\lambda)\}^{-1} \exp F(\phi(\lambda) + \epsilon(\lambda),\lambda) \\ &\leq \{m(\lambda)\epsilon(\lambda)\}^{-1} \exp F(\phi(\lambda) + \epsilon(\lambda),\lambda), \end{aligned}$$

where $0 < \theta < 1$. Thus by (9) and II) we have

$$\begin{aligned} 0 < \frac{J_3(\lambda)}{J_2(\lambda)} &\leq \frac{\exp[F(\phi(\lambda + \epsilon,\lambda) - F(\phi(\lambda),\lambda)]}{(1-\delta)\cdot m(\lambda)\cdot \epsilon(\lambda)} \left(\frac{M(\lambda)}{2\pi}\right)^{\frac12} \\ &< A\,\frac{\exp[-\frac12 \epsilon(\lambda)^2 m(\lambda)]}{(m(\lambda))^{\frac12}\cdot \epsilon(\lambda)} \to 0, \end{aligned}$$

as $\lambda \to \infty$. In a like manner we may verify III) by setting $F(u,\lambda) = t$. Hence in conclusion the function $\epsilon(\lambda)$ defined by (10) satisfies all the required conditions.

Clearly Lemma 1 is also true for the case $a > -\infty$ or $b < +\infty$, since all the inequalities occurred above still hold.

To illustrate the use of Lemma 1 we give the following simple examples:

$$1.\quad \int_0^\infty u^\lambda \exp(-u\lambda^2/(A\lambda + B\lambda^{\frac12}))\,du \sim A(eE)^{-\lambda}(2\pi/\lambda)^{\frac12}, \qquad (\lambda\to\infty),$$

$$\begin{aligned} 2.\quad &\int_0^\infty u^\lambda (1+u)^{-1} \exp(-u\lambda^2/(A\lambda + B\lambda^{\frac12}))\,du \\ &\sim (1+A)[(\phi(\lambda))^\lambda/(1+\phi(\lambda))] \exp(-\lambda^2\phi(\lambda)/(A\lambda + B\lambda^{\frac12})) \\ &\qquad \times (2\pi/(\lambda(1+A^{-1})^2 - 1))^{\frac12}, \end{aligned}$$

where A, B are arbitrary constants with $A > 0$, and

$$E = \lambda/(A\lambda + B\lambda^{\frac12}),$$

$$\phi(\lambda) = (2E)^{-1}\{[(E - 1 + 1/\lambda)^2 + 4E]^{\frac12} - E + 1 - 1/\lambda\}.$$

The expression 1. can also be verified immediately by use of the Eulerian integral and Stirling's formula. The relation 2. cannot be directly obtained by the classical Laplace asymptotic evaluation.

LEMMA 2. *If a, b are finite numbers $(a < b)$ and if in addition to hypotheses* (i), (ii), (iii) *of Lemma* 1 *we have*

(iv) $$g(u) \varepsilon L(a, b), \qquad g(\xi) \neq 0, \qquad (a < \xi < b),$$

where ξ *belongs to the Lebesgue set for* $g(u)$, *then there is a sequence of points* $\{\phi(\lambda_n)\}$ *belonging to the Lebesgue set of* $g(w)$ *with* $\lambda_n \to \infty$, *and for* $\lambda_n \to \infty$ *we have*

$$\int_a^b \exp[F(u, \lambda_n)] g(u)\, du \sim \exp[F(\phi(\lambda_n), \lambda_n)]\, g(\xi)\, (-2\pi / F_{uu}(\xi, \lambda_n))^{1/2}. \tag{11}$$

Proof. We have assumed in the hypothesis (i) that $u = \phi(\lambda)$ is a continuous function approaching $\xi (a < \xi < b)$ as $\lambda \to \infty$. Hypothesis (iv) implies that the points of the Lebesgue set for $g(u)$ are almost everywhere in (a, b). Thus by the continuity of $\phi(\lambda)$ it is easy to see that an unbounded increasing sequence $\{\lambda_n\}$ can always be chosen such that $\{\phi(\lambda_n)\}$ belong to the Lebesgue set for $g(u)$.

By use of (8) we need only show that, as $\lambda_n \to \infty$,

$$I = (-F_{uu}(\xi, \lambda_n))^{1/2} \int_a^b [g(u) - g(\xi)] \exp[F(u, \lambda_n) - F(\phi(\lambda_n), \lambda_n)]\, du = o(1).$$

Write

$$I = \int_a^{\phi(\lambda_n)-\delta} + \int_{\phi(\lambda_n)-\delta}^{\phi(\lambda_n)} + \int_{\phi(\lambda_n)}^{\phi(\lambda_n)+\delta} + \int_{\phi(\lambda_n)+\delta}^{b} = J_1 + J_2 + J_3 + J_4,$$

where $\delta > 0$ is a properly chosen fixed number. It suffices to prove that $J_3(n) + J_4(n) = o(1)$ as $n \to \infty$; for the other parts, $J_1 + J_2$, can be treated in exactly the same manner.

Recall that for every fixed large $v = \lambda_n$, the function $F(u, v)$ has a maximum at $u = \phi(\lambda_n)$, and $F(u, \lambda_n)$ is decreasing in $\phi(\lambda_n) \leq u \leq b$. Hence

$$\begin{aligned} |J_4(n)| &\leq (-F_{uu}(\xi, \lambda_n))^{1/2} \exp[F(\phi(\lambda_n) + \delta, \lambda_n) \\ &\qquad - F(\phi(\lambda_n), \lambda_n)] \int_{\phi(\lambda_n)+\delta}^{b} |g(u) - g(\xi)|\, du \\ &\leq (-F_{uu}(\xi, \lambda_n))^{1/2} \exp[\tfrac{1}{2}\delta^2 F_{uu}(\phi + \theta\delta, \lambda_n)] A, \end{aligned}$$

where $0 < \theta < 1$, and A is implied by the fact that $g \varepsilon L(a, b)$.

Clearly hypothesis (iii) implies the existence of a small number $\delta_1 > 0$ and a large number N_1 such that, e. g.

$$0 < -\tfrac{3}{4} F_{uu}(\xi, \lambda_n) < -F_{uu}(u, \lambda_n) < -\tfrac{5}{4} F_{uu}(\xi, \lambda_n), \tag{12}$$

whenever $|u - \xi| < \delta_1$, and $\lambda_n > N_1$. Since $\phi(\lambda_n) \to \xi$ we may choose M so large and δ so small that $|\phi(\lambda_n) + \theta\delta - \xi| < \delta_1$, whenever $\lambda_n > M$, where δ is fixed after choice. Thus it follows from (iii) and (12) that $-F_{uu}(\phi+\theta\delta,\lambda_n) > -\frac{3}{4}F_{uu}(\xi,\lambda_n) \to \infty$ as $\lambda_n \to \infty$, and consequently,

$$\overline{\lim_{n\to\infty}}\,|J_4(n)| \leq \lim_{n\to\infty} A(-F_{uu}(\xi,\lambda_n))^{\frac{1}{2}}\exp[\tfrac{3}{8}\delta^2 F_{uu}(\xi,\lambda_n)] = 0$$

Hence in conclusion $J_4(n) = o(1)$.

Now consider $J_3(n)$. Since ξ, $\phi(\lambda_n)$ are points belonging to the Lebesgue set for $g(u)$, we have

$$(13)\qquad \begin{aligned} &\int_\xi^u |g(x)-g(\xi)|\,dx = o(|u-\xi|), &&(|u-\xi|\to 0);\\ &\int_\phi^w |g(x)-g(\phi)|\,dx = o(|w-\phi|), &&(|w-\phi|\to 0). \end{aligned}$$

Putting $u=\phi(\lambda_n)$, $w=\xi$ in (13) and comparing, we get

$$\int_\xi^\phi |g(\phi)-g(\xi)|\,dx = o(|\phi(\lambda_n)-\xi|), \qquad (|\phi-\xi|\to 0).$$

This is equivalent to $|g(\phi)-g(\xi)| = o(1)$. That means $g(\phi(\lambda_n)) \to g(\xi)$ as $\lambda_n \to \infty$. Set

$$\alpha(u) = \int_{\phi(\lambda_n)}^u [g(x)-g(\phi(\lambda_n))]dx = o(|u-\phi(\lambda_n)|), \quad (u\to\phi(\lambda_n)),$$

and write $g(u)-g(\xi) = [g(w)-g(\phi)] + K(\phi(\lambda_n))$; so that $\alpha(\phi(\lambda_n)) \equiv 0$ and $K(\phi) = g(\phi(\lambda_n)) - g(\xi) \to 0$ as $\lambda_n \to \infty$. Then integration by parts gives

$$\begin{aligned} J_3(n) &= (-F_{uu}(\xi,\lambda_n))^{\frac{1}{2}} \int_\phi^{\phi+\delta} \{K(\phi) + [g(u)-g(\phi)]\}\exp\{F(u,\lambda_n)-F(\phi,\lambda_n)\}\\ &= (-F_{uu}(\xi,\lambda_n))^{\frac{1}{2}} \int_\phi^{\phi+\delta} K(\phi)\exp\{F(u,\lambda_n)-F(\phi,\lambda_n)\}du\\ &\quad + (-F_{uu}(\xi,\lambda_n))^{\frac{1}{2}}\alpha(\phi+\delta)\exp\{F(\phi+\delta,\lambda_n)-F(\phi,\lambda_n)\}\\ &\quad -(-F_{uu}(\xi,\lambda_n))^{\frac{1}{2}} \int_\phi^{\phi+\delta} \alpha(u)\exp\{F(u,\lambda_n)-F(\phi,\lambda_n)\}F_u(u,\lambda_n)du\\ &= J_5(n) + J_6(n) + J_7(n). \end{aligned}$$

Here $J_6(n) = o(1)$ $(n\to\infty)$ is easily inferred by the same argument as used in proving $J_4(n) = o(1)$. For proof of $J_5(n) = o(1)$ and $J_7(n) = o(1)$, we write

$$(14)\qquad \begin{aligned} F(u,\lambda_n) - F(\phi,\lambda_n) &= \tfrac{1}{2}(u-\phi)^2 F_{uu}(\xi',\lambda_n),\\ F_u(u,\lambda_n) &= (u-\phi)F_{uu}(\xi'',\lambda_n), \end{aligned}$$

where $\xi' = \phi(\lambda_n) + \theta'(u-\phi(\lambda_n))$, $\xi'' = \phi(\lambda_n) + \theta''(u-\phi(\lambda_n))$, $0 < \theta'$,

$\theta'' < 1$, $\phi \leq u \leq \phi + \delta$. Clearly $\xi' \cong \xi + \theta'(u - \xi)$, $\xi'' \cong \xi + \theta''(u - \xi)$ as $\lambda_n \to \infty$, so that by (iii) or by the same reasoning of (12) we have, for λ_n sufficiently large and for certain fixed small $\delta > 0$,

$$(15) \qquad -F_{uu}(\xi', \lambda_n) > -A_1 F_{uu}(\xi, \lambda_n),$$
$$-A_2 F_{uu}(\phi, \lambda_n) > -F_{uu}(\xi'', \lambda_n) > -A_3 F_{uu}(\phi, \lambda_n),$$

where A_1, A_2, A_3 are certain positive numbers. Now for any given $\epsilon > 0$ we can determine a large number M' and a small positive number η $(\leq \delta)$ such that

$$(16) \qquad |\alpha(u)| < \epsilon\, |u - \phi|, \qquad |K(\phi)| < \epsilon,$$

for $|u - \phi(\lambda_n)| < \eta$ and $\lambda_n > M'$. It is obvious that $J_4(n) = o(1)$, $J_6(n) = o(1)$ are also true when δ is replaced by η. Thus using (14), (15), (16) and replacing δ by η, we get

$$|J_5(n)| < \epsilon \int_\phi^{\phi+\eta} (-F_{uu}((\xi, \lambda_n))^{\frac{1}{2}} \exp\{F(u, \lambda_n) - F(\phi, \lambda_n)\}\, du$$
$$< \epsilon \int_\phi^{\phi+\eta} (-F_{uu}(\xi, \lambda_n))^{\frac{1}{2}} \exp\{\tfrac{1}{2} A_1 (u - \phi)^2 F_{uu}(\xi, \lambda_n)\}\, du$$
$$\leq \epsilon \int_0^\infty \exp\{-\tfrac{1}{2} A_1 t^2\}\, dt = A\epsilon,$$

with $t = (-F_{uu}(\xi, \lambda_n))^{1/2}(u - \phi)$, and similarly,

$$|J_7(n)| < \epsilon(-F_{uu}(\xi, \lambda_n))^{\frac{1}{2}} \int_\phi^{\phi+\eta} (u - \phi)^2\, |F_{uu}(\xi'', \lambda_n)| \exp\{\tfrac{1}{2}(u - \phi)^2 F_{uu}(\xi'', \lambda_n)\}\, du$$
$$< A\epsilon \int_\phi^{\phi+\eta} (-F_{uu}(\phi, \lambda_n))^{3/2} (w - \phi(\lambda_n))^2 \exp\{\tfrac{1}{2} A_3 (u - \phi)^2 F_{uu}(\phi, \lambda_n)\}\, du$$
$$\leq A\epsilon \int_0^\infty s^2 \exp\{-\tfrac{1}{2} A_3 s^2\}\, ds = A\epsilon,$$

with $s = (-F_{uu}(\phi, \lambda_n))^{\frac{1}{2}}(u - \phi(\lambda_n))$. Hence it follows that $\overline{\lim}\, |J_5(n)| \leq A\epsilon$, $\overline{\lim}\, |J_7(n)| \leq A\epsilon$ where $n \to \infty$. Since ϵ is arbitrary we get $J_5(n) = o(1)$, $J_7(n) = o(1)$ $(n \to \infty)$. This completes the proof of $J_3(n) + J_4(n) = o(1)$. By a similar procedure it can be shown that $J_1(n) + J_2(n) = o(1)$. Therefore our Lemma 2 is true.

LEMMA 3. *Let $F(u, v)$ satisfy all hypotheses of Lemma 1 with $a = 0$, $b = \infty$, $0 < \xi < \infty$, and let $g(u) \in L(0, \infty)$ with $g(\xi) \neq 0$, ξ being a point of the Lebesgue set for $g(u)$. Then we have*

$$(17) \qquad \int_0^\infty \exp[F(u, \lambda_n)]\, g(u)\, du$$
$$\sim \exp[F(\phi(\lambda_n), \lambda_n)]\, g(\xi)\, (-2\pi/F_{uu}(\xi, \lambda_n))^{\frac{1}{2}},$$

where $\{\lambda_n\}$ *is an unbounded increasing sequence for which* $\{\phi(\lambda_n)\}$ *belong to the Lebesgue set for* $g(u)$. *Moreover if* $g(\xi) = 0$, *then*

$$\int_0^\infty \exp[F(u,\lambda_n)]g(u)\,du \\ = o(\exp[F(\phi(\lambda_n),\lambda_n)](-2\pi/F_{uu}(\xi,\lambda_n))^{\frac{1}{2}}. \tag{18}$$

Proof. By use of Lemma 2 we need only to show that for any fixed number $\delta > 0$, we have, for $n \to \infty$,

$$I(n) = (F_{uu}(\xi,\lambda_n)/(-2\pi))^{\frac{1}{2}} \int_{\xi+\delta}^{x} |g(u)| \exp\{F(u,\lambda_n) - F(\phi(\lambda_n),\lambda_n)\}du = o(1).$$

Set $\alpha(x) = \int_{\xi+\delta}^{x} |g(u)|\,du$, so that $\alpha(\xi+\delta) = 0$, and $\alpha(\infty)$ is a finite value. Integration by parts gives

$$\begin{aligned} I(n) &= (F_{uu}(\xi,\lambda_n)/(-2\pi))^{\frac{1}{2}}\alpha(\infty)\exp\{F(\infty,\lambda_n) - F(\phi,\lambda_n)\} \\ &\quad - (F_{uu}(\xi,\lambda_n)/(-2\pi))^{\frac{1}{2}} \int_{\xi+\delta}^{\infty} \exp\{F(u,\lambda_n) - F(\phi,\lambda_n)\}\alpha(u)F_u(u,\lambda_n)du \\ &= J_1(n) - J_2(n), \end{aligned}$$

where

$$F(\infty,\lambda_n) - F(\phi,\lambda_n) = \int_{\phi(\lambda_n)}^{\infty} F_u(u,\lambda_n)\,du. \tag{19}$$

Obviously $F_u(u,\lambda_n) \leq 0$ is a monotonic decreasing function of u in (ϕ,∞) for sufficiently large λ_n, so that in case the integral of (19) does not converge we must have $F(\infty,\lambda_n) = -\infty$. We shall now prove that $J_1(n) = o(1)$, $J_2(n) = o(1)$ $(n \to \infty)$ without assuming $F(\infty,\lambda_n) = -\infty$. Since $\alpha(u)$ is bounded we have

$$\begin{aligned} |J_2(n)| &\leq A(F_{uu}(\xi,\lambda_n)/(-2\pi))^{\frac{1}{2}} \left| \int_{\xi+\delta}^{\infty} \exp\{F(u,\lambda_n) - F(\phi,\lambda_n)\}F_u(u,\lambda_n)du \right| \\ &= A(F_{uu}(\xi,\lambda_n)/(-2\pi))^{\frac{1}{2}} |\exp\{F(\infty,\lambda_n) \\ &\qquad - F(\phi,\lambda_n)\} - \exp\{F(\xi+\delta,\lambda_n) - F(\phi,\lambda_n)\}|. \end{aligned}$$

Therefore in order to establish $J_1(n) = o(1)$ and $J_2(n) = o(1)$ we need only to show that, as $n \to \infty$,

$$\lim (F_{uu}(\xi,\lambda_n)/(-2\pi))^{\frac{1}{2}} \exp\{F(\xi+\delta,\lambda_n) - F(\phi,\lambda_n)\} \to 0, \tag{20}$$

$$\lim (F_{uu}(\xi,\lambda_n)/(-2\pi))^{\frac{1}{2}} \exp\{F(\infty,\lambda_n) - F(\phi,\lambda_n)\} \to 0. \tag{21}$$

Note that $F(\xi+\delta,\lambda_n) - F(\phi,\lambda_n) = \frac{1}{2}\Delta^2 F_{uu}(\phi+\theta\Delta,\lambda_n)$, where $o < \theta < 1$ and $\Delta = \xi + \delta - \phi(\lambda_n) \to \delta > 0$. Thus by choosing suitable fixed $\delta > 0$ and using the same argument as previously employed in proving $J_4(n) = o(1)$

(see Proof of Lemma 2), we have, when $\lambda_n \to \infty$, $-F_{uu}(\phi + \theta\Delta, \lambda_n) > -\frac{3}{4}F_{uu}(\xi, \lambda_n) \to \infty$, and consequently,

$$(F_{uu}(\xi, \lambda_n)/(-2\pi))^{\frac{1}{2}}\exp\{\tfrac{1}{2}\Delta^2 F_{uu}(\phi + \theta\Delta, \lambda_n)\} = o(1).$$

This establishes (20).

To prove (16), let us notice that

$$F_u(u, \lambda_n) = \int_{\phi(\lambda_n)}^{u} F_{uu}(t, \lambda_n)\,dt < 0,$$

whenever $u > \phi(\lambda_n)$, using hypothesis (ii). Thus for n so large that that $\xi + \delta \geq \phi(\lambda_n)$ we have

$$F(\infty, \lambda_n) = F(\xi + \delta, \lambda_n) + \int_{\xi+\delta}^{\infty} F_u(u, \lambda_n)\,du < F(\xi + \delta, \lambda_n).$$

This shows clearly that (21) is implied by (20). Hence in conclusion $I(n) = o(1)$, and the formula (17) follows from Lemma 2. Moreover if we take $g(u) = g_1(u) + g_2(u)$ with $g_1(\xi) \neq 0$, $g_2(\xi) = 0$, then an application of (17) yields the equation (18) with $g(u) = g_2(w)$. Our lemma is thus completely proved.

Example. Evidently $(1 + u^{\alpha})^{-1} \varepsilon L(0, \infty)$ with $\alpha > 1$. Thus by Example 1. of Lemma 1 and by Lemma 3 we have, as $\lambda \to \infty$,

$$3. \qquad \int_0^{\infty} u^{\lambda}/(1 + u^{\alpha})\exp(-u\lambda^2/(A\lambda + \mathrm{B}\lambda^{\frac{1}{2}}))\,du$$
$$\sim (A/(1 + A^{\alpha}))(A + B/\lambda^{\frac{1}{2}})^{\lambda} e^{-\lambda}(2\pi/\lambda)^{\frac{1}{2}}.$$

It is noteworthy that the above asymptotic relation cannot be directly obtained from (8) of Lemma 1, as the integrand function does not satisfy conditions (i), (ii), (iii). If $\alpha = 1$, then Lemma 1 is applicable and we have Example 2.

Clearly Lemmas 2 and 3 are extensions of the classical Laplace asymptotic evaluation for an integral ([4; p. 78], [1; pp. 277-278]).

3. A general inversion theorem. Using the results obtained in § 2 we may now state sufficient conditions for the validity of the general inversion formula (5). For convenience we introduce the following definitions:

DEFINITION 1. *A function $\psi(u, v, t)$ is called a regular transform function with respect to a certain fixed $t = t_1 > 0$, if $\log \psi(u, v, t_1) = F(u, v, t_1)$, $F_u(u, v, t_1)$, $F_{uu}(u, v, t_1)$ are real continuous functions $(u \geq 0,\ v > 0)$ satisfying the following conditions:*

i) $F_u(u, v, t_1) = 0$ *has a continuous solution* $u = \phi(v, t_1)$ *with* $\lim_{v \to \infty} \phi(v, t_1) = \theta(t_1) > 0$,

ii) $F_{uu}(u, v, t_1) < 0$ *for all* $u > 0$ *and large* v,

iii) $F_{uu}(u, v, t_1) \sim F_{uu}(\theta(t_1), v, t_1) \to -\infty$ *as* $u \to \theta(t_1)$, $v \to \infty$.

A simple example of a regular transform function is

$$\psi(u, v, t) = u^v \exp\{-uv^2/(vt + v^{\frac{1}{2}} \sin t)\}.$$

Of course we may construct as many such transform functions as we like.

DEFINITION 2. $\lim \phi(v, t) = \theta(t)$ *is called a limiting solution of the equation* $(\partial/\partial u) \log \psi(u, v, t) = 0$.

Clearly for every fixed $t > 0$, the functions $F(u, v, t)$, $F_{uu}(u, v, t)$, $F_{uu}(u, v, t)$ etc. can be considered as functions of the two variables u, v. Thus by Lemmas 1 and 3 we easily establish the following theorem.

THEOREM 1. *If* $\psi(u, v, t) = \exp F(u, v, t)$ *is a regular transform function with respect to a certain* $t > 0$, *if* $g(u) \varepsilon L(0, \infty)$, *and if* $f(\lambda, t)$ *is defined by* (4), *then*

$$g(\theta(t)) = \lim_{\lambda \to \infty} \left(\frac{-F_{uu}(u, \lambda, t)}{2\pi(\psi(u, \lambda, t))^2} \right)^{\frac{1}{2}}_{u=\phi(\lambda, t)} f(\lambda, t), \tag{5 bis}$$

where λ *tends to* ∞ *through such a sequence* $\{\lambda_n\}$ *that* $\{\phi(\lambda_n, t)\}$ *together with its limit* $\lim \phi = \theta(t)$ *belong to the Lebesgue set for* $g(u)$.

Proof. Clearly in the formulas (17) and (18) we may replace $F_{uu}(\xi, \lambda_n)$ by $F_{uu}(\phi(\lambda_n), \lambda_n)$. Let $F(u, v) = F(u, v, t)$ so that $\exp F(u, v) = \psi(u, v, t)$, where t is fixed. Then from (17) we may infer that, as $\lambda_n \to \infty$,

$$\frac{(-F_{uu}(\phi(\lambda_n, t), \lambda_n, t))^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}} \exp F(\phi(\lambda_n, t), \lambda_n, t)} \int_0^\infty \psi(u, \lambda_n, t) g(u) du \to g(\theta(t)),$$

where $\phi(\lambda_n, t) = \phi(\lambda_n)$, $\theta(t) = \xi$, $g(\xi) \neq 0$. This is just equivalent to (5 bis) with $g(\theta(t)) \neq 0$. Moreover if $g(\theta(t)) = 0$, then by (18) we have, for $\lambda_n \to \infty$,

$$\frac{(-F_{uu}(\phi(\lambda_n, t), \lambda_n, t))^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}} \exp F(\phi(\lambda_n, t), \lambda_n, t)} \int_0^\infty \psi(u, \lambda_n, t) g(u) du = o(1).$$

Hence in either case the formula (5 bis) is true.

COROLLARY. *If* $\psi(u, v, t)$ *is a regular transform function and if* $g(u)$

is a continuous function, absolutely integrable over $(0, \infty)$, *then* (5 bis) *is valid, where* λ *tends to* ∞ *through any sequence.*

4. Specialization. We now consider transform functions of the form

$$(22) \qquad \psi(u, \lambda, t) = \exp\{\lambda \Phi(u, t)\}, \qquad (u \geq 0, t > 0),$$

where $\Phi(u, t)$ is a real valued function. If $\Phi_u(u, t)$, $\Phi_{uu}(u, t)$ are continuous functions such that

i)^1 there is a continuous function $u = \xi(t) > 0$ satisfying the equation $\Phi_u(\xi(t), t) = 0$.

ii)^1 $\Phi_{uu}(u, t) < 0$ for all $u, t > 0$.

then clearly $\psi(u, \lambda, t)$ is a regular transform function, because the condition iii) of Definition 1 of § **3** is automatically satisfied. Obviously in the present case our formula (5 bis) takes the form

$$(5)^1 \qquad g(\xi(t)) = \lim_{\lambda \to \infty} \frac{[-\lambda \Phi_{uu}(\xi(t), t)]^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}} \exp[\lambda \Phi(\xi(t), t)]} f(\lambda, t),$$

where $f(\lambda, t)$ is defined by (4), and $\lambda \to \infty$ through any real sequence.

As another particular application of Theorem 1, we take

$$(23) \qquad \psi(u, \lambda, t) = \exp\{-\lambda u \beta(\lambda, t)\} u^{\lambda},$$

where $\beta(\lambda, t)$ is an arbitrary continuous function with $\beta(\lambda, t) > 0$, $\lim \beta(\lambda, t) > 0$ $(\lambda \to \infty)$. Obviously we have

$$\int_0^{\infty} g(u) \exp\{-\lambda u \beta(\lambda, t)\} u^{\lambda} \, du = (-1)^{\lambda} f^{(\lambda)}(\lambda \beta(\lambda, t)),$$

where $f(x)$ is the Laplace transform of $g(u)$ (cf. (3)). Thus using (5) or (5 bis) we easily obtain

$$(24) \qquad g(\alpha(t)) = \lim_{\lambda \to \infty} (-1)^{\lambda} f^{(\lambda)}(\lambda \beta(\lambda, t)) (\lambda \beta(\lambda, t))^{\lambda+1} / \lambda!,$$

where $g(u) \varepsilon L(0, \infty)$, $\alpha(t) > 0$, $\lim \alpha(t) \beta(\lambda, t) = 1$ $(\lambda \to \infty)$ and $\lambda \to \infty$ through any sequence such that $\{\beta(\lambda, t)^{-1}\}$ belongs to the Lebesgue set for $g(u)$.

Recall that the continuity condition imposed on $\phi(\lambda, t) = \beta(\lambda, t)^{-1}$ is only used for asserting the existence of $\{\phi(\lambda_n, t)\}_{n=1}^{\infty}$ (see Proof of Lemma 2), so that it may be replaced by a direct assumption on $\{\beta(\lambda_n, t)^{-1}\}_{n=1}^{\infty}$. Thus by taking $\alpha(t) = t$, $\beta(\lambda, t) = 1/\phi(t)$ in the equation (24), we have

THEOREM 2. *Let $f(x)$ be the Laplace transform of $g(u)$ with $g(u) \varepsilon L(0, \infty)$, and let a sequence of points $\{\phi(n)\}$ together with its limit $\lim \phi(n) = t > 0$ belong to the Lebesgue set for $g(u)$. Then*

$$g(t) = \lim_{n\to\infty} (-1)^n f^{(n)}(n/\phi(n))(n/\phi(n))^{n+1}/n!. \tag{25}$$

This theorem may be regarded as an analogue of the result in Widder's book [1] (cf. Theorem 6b of [1; pp. 289-290]).

KING'S COLLEGE, ABERDEEN, SCOTLAND.

REFERENCES.

[1] D. V. Widder, *The Laplace Transform*, Princeton University Press (1946).

[2] E. L. Post, "Generalized Differentiation," *Transactions of the American Mathematical Society*, vol. 32 (1930), pp. 723-781.

[3] E. K. Haviland, "On an asymptotic expression," *Quarterly Journal of Mathematics* (Oxford), vol. 7 (1936), pp. 152-157.

[4] G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, New York (1945).

ON THE SPHERICAL CONVERGENCE OF MULTIPLE FOURIER SERIES.*

By JOSEPHINE MITCHELL.

---

1. **Introduction.** In this paper we propose to use the methods and results of the lattice point problem for the hypersphere in number theory to obtain new results for the spherical convergence of multiple Fourier series. The set of functions $\{(2\pi)^{-q}e(n_1x_1+\cdots+n_qx_q)$ $(n_k=0,\pm 1,\pm 2,\cdots;$ $k=1,\cdots,q)$ (where $e(X)=e^{iX}$ throughout this paper), forms an orthonormal system on the hypercube $C(-\pi\leq x_k\leq\pi;\ k=1,\cdots,q)$, that is,

$$(1.1)\quad (2\pi)^{-q}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}e(m_1x_1+\cdots+m_qx_q)e(-n_1x_1-\cdots-n_qx_q)\times dx_1\cdots dx_q = \prod_{k=1}^{q}\delta_{m_kn_k},$$

which is complete with respect to real Lebesgue integrable functions defined on C. We form the series

$$(1.2)\quad \sum a_{n_1\ldots n_q}e(n_1x_1+\cdots+n_qx_q)\quad (n_k=-\infty \text{ to } \infty;\ k=1,\cdots,q),$$

where $\{a_{n_1\ldots n_q}\}$ is a sequence of complex numbers. Summing (1.2) by spherical, instead of rectangular, partial sums, we consider the limit of the sequence

$$(1.3)\qquad S_R=\sum_{\nu\leq R}a_{n_1\ldots n_q}e(n_1x_1+\cdots+n_qx_q)\qquad (\nu=n_1^2+\cdots+n_q^2)$$

as $R\to\infty$, R being the sequence of integers, which can be represented as the sum of q squares of integers, so that for $q\geq 4$, R takes on all non-negative integral values [6].[1] We ask ourselves what condition must be imposed on the sequence $\{a_{n_1\ldots n_q}\}$ in order that $\lim_{R\to\infty}S_R$ exist and show that *if*[2]

$$\sum(n_1^2+\cdots+n_q^2)^p\,|a_{n_1\ldots n_q}|^2<\infty,$$

---

* Received June 5, 1950; presented to the American Mathematical Society, April 28, 1950.

[1] The numbers in brackets refer to the references at the end of the paper.

[2] The summation for this series is always $n_k=-\infty$ to $+\infty$, $k=1,\cdots,q$, and will be omitted.

$\lim_{R\to\infty} S_R$ *exists almost everywhere on* C, *where* $p = 1/3$ *for* $q = 2$ *and* $p = \frac{1}{4}q$ *for* $q \geq 4$, *while for* $q = 3$ *the sequence* $(n_1^2 + n_2^2 + n_3^2)^p$ *is replaced by* $(n_1^2 + n_2^2 + n_3^2)^{2/3}(\log(n_1^2 + n_2^2 + n_3^2)^{5/3}$. This theorem represents a considerable improvement over known results due to Chandrasekharan and Minakshisundaran, where $p = \frac{1}{2}q$ [1, 2]. We prove it by finding a bound for the Lebesgue function

$$(1.4)\quad L_R(\alpha) = L_R(\alpha_1, \cdots, \alpha_q) = \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} | K_R(\alpha - x) | \, dx_1 \cdots dx_q,$$

where

$$(1.5)\quad K_R(\alpha - x) = K_R(\alpha_1 - x_1, \cdots, \alpha_q - x_q)$$
$$= 1/(2\pi)^q \sum_{\nu \leq R} e(n_1(\alpha_1 - x_1) + \cdots + n_q(\alpha_q - x_q)),$$

and then proceeding by classical methods used in Fourier and orthogonal series to the conclusion concerning $\lim S_R$. Finer results are obtained for $q = 2$ and 3, namely, $L_R(\alpha) = O(R^{1/3})$ and $O(R^{2/3}(\log R)^{5/3})$ respectively, by using new results on exponential sums due to L. K. Hua [3] in the lattice point problem for the circle and sphere.

2. The case $q = 2$.

2. 1. We prove

THEOREM 2. 1. *The Lebesgue function* $L_R(\alpha, \beta)$ *is* $O(R^{1/3})$.

Proof. From the periodicity of the exponential system

$$L_R(\alpha, \beta) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} | K_R(x, y) | \, dx \, dy.$$

In the proof of this theorem we proceed as far as possible by means of the method used in the rectangular summation of Fourier series for the corresponding problem, but in order to complete the analysis some number-theoretic results are introduced (cf. Section 2. 3). Thus we subdivide the square $(-\pi \leq x \leq \pi, \ -\pi \leq y \leq \pi)$ into smaller rectangles, on each of which a suitable bound for $K_R(x, y)$ is utilized, by drawing lines parallel to the axes through the points x and $y = -\pi + \epsilon, -\epsilon, 0, \epsilon, \pi - \epsilon$, where ϵ is an arbitrary small positive number. The square is divided into rectangles of 3 different types, of which $(0 \leq x \leq \epsilon, 0 \leq y \leq \epsilon)$, $(0 \leq x \leq \epsilon, \epsilon \leq y \leq \pi - \epsilon)$ and $(\epsilon \leq x \leq \pi - \epsilon, \epsilon \leq y \leq \pi - \epsilon)$ are examples. (The fact that $| x - y |$

$< 2\pi$ is used in the discussion of (2.17).) Over squares of side ϵ we use the bound $|K_R(x,y)| \leq \sum_{m^2+n^2\leq R} 1 = O(R)$ [6], to obtain, for example,

$$\int_0^\epsilon \int_0^\epsilon |K_R(x,y)|\,dx\,dy = O(R\epsilon^2). \tag{2.1}$$

For rectangles of dimension ϵ times $\pi - 2\epsilon$, if $|y| \geq \epsilon$, we proceed as follows, $[z]$ being the integral part of z,

$$\begin{aligned} K_R(x,y) &= \sum_{-R^{1/2}\leq m\leq R^{1/2}} e(mx) \sum_{n=-[(R-m^2)^{1/2}]}^{[(R-m^2)^{1/2}]} e(ny) \\ &= 1/(e(y)-1) \sum_{-R^{1/2}\leq m\leq R^{1/2}} e(mx)\{e(([(R-m^2)^{\frac{1}{2}}]+1)y) \\ &\qquad - e(-[(R-m^2)^{\frac{1}{2}}]y)\} \\ &= O(R^{\frac{1}{2}} \csc \tfrac{1}{2}y). \end{aligned} \tag{2.2}$$

Hence, for example,

$$\int_\epsilon^{\pi-\epsilon} dy \int_0^\epsilon |K_R(x,y)|\,dx = O(\epsilon R^{\frac{1}{2}} \log \epsilon^{-1}). \tag{2.3}$$

Similarly

$$\begin{aligned} &\int_\epsilon^{\pi-\epsilon} dy \int_{y-\frac{1}{2}\epsilon}^{y+\frac{1}{2}\epsilon} |K_R(x,y)|\,dx \\ &\quad = O(R^{\frac{1}{2}} \int_\epsilon^{\pi-\epsilon} \csc\tfrac{1}{2}y\,dy \int_{y-\frac{1}{2}\epsilon}^{y+\frac{1}{2}\epsilon} dx) = O(R^{\frac{1}{2}}\epsilon \log \epsilon^{-1}). \end{aligned} \tag{2.4}$$

(Also, of course, $\int_\epsilon^{\pi-\epsilon} dx \int_{x-\frac{1}{2}\epsilon}^{x+\frac{1}{2}\epsilon} |K_R(x,y)|\,dy = O(R^{\frac{1}{2}}\epsilon \log \epsilon^{-1})$); a result which we shall need later. In the course of our argument we note that all bounds introduced except those involving $O(|x-y|^{-1})$ may be integrated over $(\epsilon \leq y \leq \pi - \epsilon,\ y - \frac{1}{2}\epsilon \leq x \leq y + \frac{1}{2}\epsilon)$ as well as over $(\epsilon \leqq x \leqq \pi - \epsilon,\ \epsilon \leq y \leq \pi - \epsilon)$, obtaining the same or even sharper results.

2.2. In order to discuss the sum $\sum_{m^2+n^2\leq R}$ over rectangles of the third type we subdivide the circle $m^2 + n^2 \leq R$ by the lines $n = m$, $n = -m$, $m = 0$ and $n = 0$. On the line $n = m$, for example, we have

$$\begin{aligned} K_R(x,y) &= \sum_{m=-[(\frac{1}{2}R)^{1/2}]}^{[(\frac{1}{2}R)^{1/2}]} e((x+y)m) \\ &= \{e((x+y)([(\tfrac{1}{2}R)^{\frac{1}{2}}]+1)) \\ &\qquad - e(-(x+y)[(\tfrac{1}{2}R)^{\frac{1}{2}}])\}/\{e(x+y)-1\} \end{aligned} \tag{2.5}$$

$= Q_1$. Hence

$$(2.6)\qquad \int_\epsilon^{\pi-\epsilon}\int_\epsilon^{\pi-\epsilon} |\, Q_1 \,|\, dx\, dy = O\Big(\int_\epsilon^{\pi-\epsilon}\int_\epsilon^{\pi-\epsilon} (x+y)^{-1} dx\, dy\Big)$$

$$= O\Big(\int_\epsilon^{\pi-\epsilon} \log(x+y) \Big[_{y=\epsilon}^{\pi-\epsilon} dx\Big),$$

and the second mean value theorem used on the latter term gives

$$O\Big(\log(\epsilon+y)\int_\epsilon^{x_0} dx + \log(\pi-\epsilon+y)\int_{x_0}^{\pi-\epsilon} dx \Big[_{y=\epsilon}^{\pi-\epsilon}\Big)\ (\epsilon \leq x_0 \leq \pi-\epsilon),$$

which is $O(|\log \epsilon|)$. On the other hand, for example, for the rectangle $(\epsilon \leq x \leq \pi - \epsilon, -\pi + \epsilon \leq y \leq -\epsilon)$, we use the procedure in (2.4) over $(\epsilon \leq x \leq \pi - \epsilon, -x - \tfrac{1}{2}\epsilon \leq y \leq -x + \tfrac{1}{2}\epsilon)$ and then, as is (2.6)

$$(2.7)\qquad \Big\{\int_\epsilon^{\pi-1.5\epsilon}\int_{-\pi+\epsilon}^{-x-\frac{1}{2}\epsilon}\int_{1.5\epsilon}^{\pi-\epsilon} + \int_{-x+\epsilon/2}^{-\epsilon}\Big\} |\, Q_1 \,|\, dx\, dy = O(\log \epsilon^{-1}).$$

Analogous results are obtained when $m = -n$, $m = 0$ or $n = 0$.

The remaining sums to be considered in $\sum_{m^2+n^2\leq R}$ are all of the form

$$(2.8)\qquad \sum_{0<m\leq a}\ \sum_{m<n\leq(R-m^2)^{1/2}} e(\pm mx \pm ny)$$

$$(\epsilon \leq x \leq \pi-\epsilon, \epsilon \leq y \leq \pi - \epsilon)\ (a = (\tfrac{1}{2}R)^{\frac{1}{2}}),$$

or this sum with m and n interchanged, of which, as we shall see, it is sufficient to study the cases $mx + ny$ and $-mx + ny$. Hence consider first

$$(2.9)\qquad S = \sum_{0<m\leq a}\ \sum_{m<n\leq(R-m^2)^{1/2}} e(mx+ny)$$

$$= \frac{e(y)}{e(y)-1}\sum_{0<m\leq a} e([(R-m^2)^{\frac{1}{2}}]y + mx)$$

$$- \frac{e(y)}{e(y)-1}\sum_{0<m\leq a} e((x+y)m)$$

$= S_1 + S_2$. Now

$$S_2 = (e((x+y)[a]) - 1)(e(y)-1)^{-1}(e(x+y)-1)^{-1}e(x+2y)$$

$= O(\csc \tfrac{1}{2}y \csc \tfrac{1}{2}(x+y))$, so that

$$(2.10)\qquad \int_\epsilon^{\pi-\epsilon}\int_\epsilon^{\pi-\epsilon} |\, S_2 \,|\, dx\, dy = O\Big(\iint_{2\epsilon\leq x+y\leq\pi} y^{-1}(x+y)^{-1} dx\, dy$$

$$+ \iint_{\pi\leq x+y\leq 2\pi-2\epsilon} y^{-1}(2\pi - x - y)^{-1} dx\, dy\Big)$$

$$= O\Big(\int_\epsilon^{\pi-\epsilon}\int_\epsilon^{\pi-\epsilon} (x+y)^{-1}\{y^{-1} + (\pi-y)^{-1}\} dx\, dy\Big)$$

$$= O\Big(\int_\epsilon^{\pi-\epsilon} \{y^{-1} + (\pi-1)^{-1}\}\log(x+y)\, dy \Big[_{x=\epsilon}^{\pi-\epsilon}\Big)$$

and using the second mean value theorem on $\log(x+y)$, we get (2.10) to be $O(\log^2 \epsilon^{-1})$.

Since the exponent $[(R-m^2)^{\frac12}]$ in S_1 is discontinuous, this sum cannot be dealt with by standard methods. Consequently we replace $[(R-m^2)^{\frac12}]$ by $(R-m^2)^{\frac12}-\frac12$ and consider

$$(2.11) \qquad S_{11}=\frac{e(y)}{e(y)-1}\sum_{0<m\leq a} e(mx+y(R-m^2)^{\frac12}-\tfrac12 y)$$

and

$$(2.12) \qquad S_{12}=\frac{e(y)}{e(y)-1}\sum_{0<m\leq a} \{e(mx+y[(R-m^2)^{\frac12}]) - e(mx+y(R-m^2)^{\frac12}-\tfrac12 y)\},$$

where $S_{11}+S_{12}=S_1$. The exponent $2\pi f(u)=ux+y(R-u^2)^{\frac12}$ in S_{11} satisfies the hypotheses of the following theorem which is due to van der Corput [8]:

Let $D=\int_0^a e(2\pi f(u))\,du-\sum_{0<m\leq a} e(2\pi f(m))$. *If* $f'(u)$ *is monotonic and* $|f'(u)|\leq\frac12$, *then* $D=O(1)$. Hence

$$(2.13) \qquad S_{11}(e(y)-1)e(-\tfrac12 y)=\int_0^a e(ux+y(R-u^2)^{\frac12})\,du+O(1) = I_1+O(1).$$

Setting $rv=xu+y(R-u^2)^{\frac12}$ in I_1, where $r=(x^2+y^2)^{\frac12}$ and $ru=xv-y(R-v^2)^{\frac12}$, we get

$$(2.14) \qquad I_1=r^{-1}\int_{v_1}^{v_2}(x+yv/(R-v^2)^{\frac12})e(rv)\,dv=I_{11}+I_{12}$$

$(v_1=yR^{\frac12}r^{-1},\ v_2=(x+y)ar^{-1})$. Now $I_{11}=O(x(x^2+y^2)^{-1})$ so, that as in (2.10)

$$(2.15) \qquad \int_\epsilon^\pi\int_\epsilon^\pi \csc\tfrac12 y\,|I_{11}|\,dx\,dy=O(\log^2\epsilon^{-1}).$$

The function $g(v)=v(R-v^2)^{-\frac12}$ in the integral I_{12} is montone and $g(v_1)=yx^{-1}$, $g(v_2)=(x+y)|x-y|^{-1}$ so that by the second mean value theorem $I_{12}=O(y^2x^{-1}(x^2+y^2)^{-1})+O(y(x+y)|x-y|^{-1}(x^2+y^2)^{-1})$. As for (2.15), $\int_\epsilon^\pi\int_\epsilon^\pi\csc\frac12 y\,O(y^2x^{-1}(x^2+y^2)^{-1})\,dx\,dy=O(\log^2\epsilon^{-1})$, and, using (2.4), with

$$(2.16) \qquad \left\{\int_\epsilon^{\pi-\frac12\epsilon}\int_{y+\frac12\epsilon}^{\pi}+\int_{1.5\epsilon}^{\pi}\int_\epsilon^{y-\frac12\epsilon}\right\}\csc\tfrac12 y\,y(x+y)|x-y|^{-1}(x^2+y^2)^{-1}dx\,dy = O(\log^2\epsilon^{-1}),$$

we see that both $\int_\epsilon^\pi \int_\epsilon^\pi \csc \frac{1}{2}y \mid I_{12} \mid dx\, dy$ and $\int_\epsilon^\pi \int_\epsilon^\pi \mid S_{11} \mid dx\, dy$ are $O(\log^2 \epsilon^{-1})$ for all $\mid x - y \mid \geq \frac{1}{2}\epsilon$.

Before considering the sum S_{12} of (2.12) we discuss that sum of (2.8) whose exponent is $-mx + ny$. In particular we need only consider the sum analogous to S_{11}, namely,

$$T_{11} = \frac{e(y)}{e(y) - 1} \sum_{0 < m \leq a} e(-mx + y(R - m^2)^{\frac{1}{2}} - \tfrac{1}{2}y), \tag{2.17}$$

since sums corresponding to S_2 and S_{12} may be treated by the same methods as used for these sums. Here although the function $2\pi f(u) = -ux + y(R - u^2)^{\frac{1}{2}}$ has a monotone derivative, it is not bounded by $\frac{1}{2}$, indeed $f'(0) = -x(2\pi)^{-1}$, $f'(a) = -(x + y)(2\pi)^{-1}$ so that for $(\epsilon \leq x \leq \pi - \epsilon, \epsilon \leq y \leq \pi - \epsilon)$, $\epsilon(2\pi)^{-1} \leq \mid f'(u) \mid \leq 1 - \epsilon\pi^{-1}$, and hence the hypotheses of the van der Corput theorem do not hold. However the same method of proof is applicable. By the Euler summation formula [5]

$$\begin{aligned}(e(y) - 1)e(-\tfrac{1}{2}y)T_{11} = \int_0^a & e(-ux + y(R - u^2)^{\frac{1}{2}})\,du \\ & - e(a(-x + y))b(a) \\ - \tfrac{1}{2}e(yR^{\frac{1}{2}}) + \int_0^a & b(u)\,d(e(-ux + y(R - u^2)^{\frac{1}{2}})),\end{aligned} \tag{2.18}$$

where $b(u) = u - [u] - \frac{1}{2} = O(1)$ has the series representation

$$-\sum_{\nu=1}^{\infty} \sin 2\pi\nu/\pi\nu \quad ([u] < u < [u] + 1),$$

which has uniformly bounded partial sums. The two middle terms of the right side of (2.18) are $O(1)$, so that the integral of their absolute value, multiplied by $\csc \frac{1}{2}y$, over $(\epsilon \leq x \leq \pi - \epsilon, \epsilon \leq y \leq \pi - \epsilon)$ contributes $O(\log \epsilon^{-1})$. Now consider the term

$$\begin{aligned}\int_0^a & \{e(2\pi f(u) + \nu u) - e(2\pi f(u) - \nu u)\}f'(u)\,du \\ & = (2\pi i)^{-1} \int_0^a f'(u)(f'(u) + \nu)^{-1} d(e(2\pi f(u) + \nu u)) \\ & \quad - (2\pi i)^{-1} \int_0^a f'(u)(f'(u) - \nu)^{-1} d(e(2\pi f(w) - \nu u)).\end{aligned} \tag{2.19}$$

The functions $g_1(u) = f'(\mathrm{u})(f'(u) + \nu)^{-1}$ and $g_2(u) = f'(u)(f'(u) - \nu)^{-1}$ are monotone and do not change sign in $(0, a)$. Also, $\mid g_1(u) \mid \leq \nu^{-1}$ for $\nu \geq 1$ and $\mid g_2(u) \mid \leq (\nu - 1)^{-1}$ for $\nu \geq 2$, while for $\nu = 1$, $\mid g_2(0) \mid = \mid x(2\pi - x)^{-1} \mid \leq 1$, $\mid g_2(a) \mid = (x + y)(2\pi - x - y)^{-1} \leq \pi\epsilon^{-1}$. Using the second mean value theorem on (2.19), we thus get

$$(2.20)\qquad J_1 = \int_0^a b(u)\,d(e(-ux + y(R-u^2)^{\frac{1}{2}})) \\ = O(1) + O((x+y)(2\pi - x - y)^{-1}),$$

so that

$$(2.21)\qquad \int_\epsilon^{\pi-\epsilon}\int_\epsilon^{\pi-\epsilon} \csc \tfrac{1}{2}y \mid J_1 \mid dx\,dy = O(\log \epsilon^{-1}) \\ + O(\int_\epsilon^{\pi-\epsilon} \log(2\pi - x - y)y^{-1}dy \Big[_{x=\epsilon}^{\pi-\epsilon}) \\ = O(\log^2 \epsilon^{-1}).$$

Similarly in the first integral of (2.18), since $f'(u)$ is monotone and different from zero in $(0, a)$, $(f'(u))^{-1}$ is monotone and bounded in $(0, a)$ and using the second mean value theorem we get

$$J_2 = \int_0^a e(2\pi f(u))\,du = -2\pi x^{-1}O(1) - 2\pi(x+y)^{-1}O(1).$$

so that

$$\int_\epsilon^{\pi-\epsilon}\int_\epsilon^{\pi-\epsilon} \csc \tfrac{1}{2}y \mid J_2 \mid dx\,dy = O(\log^2 \epsilon^{-1})$$

as before. Thus finally

$$\int_\epsilon^{\pi-\epsilon}\int_\epsilon^{\pi-\epsilon} \mid T_{11} \mid dx\,dy = O(\log^2 \epsilon^{-1}),$$

2. 3. There now remains to consider the sum S_{12} of (2.12). In this case we have been unsuccessful in using a direct analysis,—the difficulty apparently arising from the discontinuity of $[(R-m^2)^{\frac{1}{2}}]$; indeed the only property of the exponential function that we use is that of boundedness. Since $e(y[(R-m^2)^{\frac{1}{2}}]) - e(y(R-m^2)^{\frac{1}{2}} - \frac{1}{2}y) = yb((R-m^2)^{\frac{1}{2}})O(1)$,

$$(2.22)\qquad S_{12} = O(y \csc \tfrac{1}{2}y \sum_{0<m<a} \mid b((R-m^2)^{\frac{1}{2}}) \mid).$$

We now use the following theorem to be found in [3, Ch. XII]:

Let $f(w)$ *be a function with two continuous derivatives in* $1 \leq u \leq P$, *and* $0 < f'(u) < 1$, $f''(u) > z^{-3}$. *Then* $\sum_{u=1}^{P} \{f(u)\} = \frac{1}{2}P + O(z^2)$, *where* $\{f(u)\} = \min(f(u) - [f(u)], [f(u)] + 1 - f(u))$.

Taking $f(u) = 1 - (R-u^2)^{\frac{1}{2}}$, we find that both $f'(u)$ and $f''(u)$ are monotonic non-decreasing in $(0, a)$ and positive. Also,

$$f'(0) = 0 < f'(u) < f'(a) = 1, \qquad f''(u) > f''(0) = R^{-\frac{1}{2}}$$

and $z = R^{1/6}$. Thus all the hypotheses of the theorem are satisfied. Now

$\{\xi\} = \{\xi - 1\}$ and $\{\xi\} = \{-\xi\}$ so that $\{f(u)\} = \{(R - u^2)^{\frac{1}{2}}\}$. Also, $0 \geq b(\xi) = \{\xi\} - \frac{1}{2}$ for $[\xi] \leq \xi \leq [\xi] + \frac{1}{2}$, and $0 \leq b(\xi) = \frac{1}{2} - \{\xi\}$ for $[\xi] + \frac{1}{2} \leq \xi \leq [\xi] + 1$, that is $|\, b(\xi) \,| = \frac{1}{2} - \{\xi\}$. Hence

$$S_{12} = O(y \csc \tfrac{1}{2}y \sum_{0<m<a} (\tfrac{1}{2} - \{(R - m^2)^{\frac{1}{2}}\})) \\ = O(R^{1/3}), \tag{2.23}$$

and thus $\int_\epsilon^\pi \int_\epsilon^\pi |\, S_{12} \,| \, dx\, dy = O(R^{1/3})$. Furthermore this result is independent of the signs used in the term $e(+ mx + ny)$, and the proof of Theorem 2.1 is complete upon choosing $\epsilon = R^{-\frac{1}{3}}$.

2.4. THEOREM 2.2. *If* $\sum (m^2 + n^2)^{1/3} |\, a_{mn} \,|^2 < \infty$, *then* $\lim_{R\to\infty} S_R$ *exists almost everywhere on* $C(-\pi \leq x \leq \pi, -\pi \leq y \leq \pi)$, *where*

$$S_R\ (\alpha, \beta) = \sum_{\nu \leq R} \sum_{\nu = m^2+n^2} a_{mn} e(m\alpha + n\beta).$$

Following a classical method of proof of Fourier and orthogonal series [4], we first show by means of Theorem 2.1 that S_R is $O(R^{1/6})$ a. e. (almost everywhere) on C, then that the convergence of $\sum (m^2 + n^2)^{1/3} |\, a_{mn} \,|^2$ implies the convergence of a certain subsequence of $\{S_R\}$ and finally that these two results imply the convergence of the sequence $\{S_R\}$ a. e. on C.

LEMMA 2.1. *If* $\sum |\, a_{mn} \,|^2 < \infty$, *then* $S_R(\alpha, \beta) = O(R^{1/6})$ *a. e. on* C.

Proof. Since $\sum |\, a_{mn} \,|^2 < \infty$, by the Riesz-Fischer Theorem extended to more variables there exists a real Lebesgue square integrable function f, whose Fourier coefficients, $(4\pi)^{-2} \int\int_C f(x, y) e(-(mx + ny))\, dx\, dy$, equal a_{mn}.

Now following a proof to be found in [4], we set $v_R = \max_{0<\nu\leq R} S_\nu(P) \nu^{-1/6} = t^{-1/6} S_t(P)$ $(R \neq 0)$, $v_0 = a_{00}$, where $P = (\alpha, \beta)$ and $t = t(P)$. The sequence $\{v_R\}$ $(R \neq 0)$ is monotone non-decreasing and we show that $I_R = \int\int_C v_R dx\, dy = O(1)$. The integral

$$I = \int\int_C f(P')\, dx'dy' \int\int_C t^{-1/6} K_t(P - P')\, dx\, dy,$$

where $P - P' = (x - x', y - y')$, and the Schwarz inequality, setting $B = \int\int f^2 dx\, dy$, we get $I^2_R \leq B \int\int |\, \int\int t^{-1/6} K_t(P - P')\, dx\, dy \,|^2\, dx'dy'$. Evaluating the latter integral [cf. 8, p. 253], we have

$$\int\int dx'dy' \int\int t^{-1/6} K_t(P - P')\, dx\, dy \int\int t_0^{-1/6} K_{t_0}(P_0 - P')\, dx_0 dy_0 \\ = \int\int\int\int t^{-1/6} t_0^{-1/6} K_T(P - P')\, dx\, dy\, dx_0 dy_0 \quad (T = \min(t, t_0)),$$

since by the orthonormal relations $\int\int K_t(P-P')K_{t_0}(P_0-P')\,dx'dy' = K_T(P-P_0)$. Consequently

$$\begin{aligned} I^2_R &\leq B \int\int\int\int (tt_0)^{-1/6}\, | K_T(P-P_0) |\, dx\, dy\, dx_0 dy_0 \\ &\leq B\{\int\int t^{-1/3} dx\, dy \int\int | K_t(P-P_0) |\, dx_0 dy_0 \\ &\qquad + \int\int t_0^{-1/3} dx_0 dy_0 \int\int | K_{t_0}(P-P_0) |\, dx\, dy, \end{aligned}$$

which by Theorem 2.1 is $O(1)$. Therefore $\lim_{R\to\infty} v_R(P)$ exists a. e. and the sequence $\{S_R(P)/R^{1/6}\}$ is bounded above a. e. Similarly the sequence is bounded below a. e. and the lemma is proved.

LEMMA 2.2. *If the series* $\sum \alpha_\nu (\alpha_\nu > 0)$ *converges, then there exists a monotone sequence* $\{p_\nu\}, p_\nu \to \infty$, *such that* $\sum p_\nu \alpha_\nu < \infty$.

This lemma is well-known.

LEMMA 2.3. *If* $\sum | a_{mn} |^2 (m^2+n^2)^{1/3} < \infty$, *then* $\{S_{\nu_k}\}$ *approaches a limit a. e. on* C, *where* $\{\nu_k\}$ *is a subsequence of the sequence* $\{R\}$ *of* (1.3) *satisfying* $k^3 \leq \nu_k < (k+1)^3$.

Proof. By the orthonormal properties we have

$$\int\int | f(P) - S_{\nu_k} |^2 dx\, dy = \sum_{\nu < \nu_k} \sum_{\nu = m^2+n^2} | a_{mn} |^2 = r_{\nu_k}.$$

The series $\sum_{k=0}^{\infty} r_{\nu_k}$ is convergent, for by Abel's transformation we get

$$\begin{aligned} \sum_{k=0}^{\infty} r_{\nu_k} &= \sum_{k=0}^{\infty} k(r_{\nu_k} - r_{\nu_{k+1}}) + \lim_{k\to\infty} k r_{\nu_{k+1}} \\ &= \sum_{k=0}^{\infty} k \sum_{\nu_k < \nu \leq \nu_{k+1}} \sum_{\nu=m^2+n^2} | a_{mn} |^2 + \lim_{k\to\infty} k \sum_{\nu > \nu_{k+1}} \sum_{\nu = m^2+n^2} | a_{mn} |^2. \end{aligned}$$

But $k \leq \nu_k^{1/3} < \nu^{1/3} = (m^2+n^2)^{1/3}$. Hence the right hand side is less than or equal to

$$\sum_{k=0}^{\infty} \sum_{\nu_k < \nu \leq \nu_{k+1}} \sum_{\nu = m^2+n^2} (m^2+n^2)^{1/3} | a_{mn} |^2 + \lim_{k\to\infty} \sum_{\nu > \nu_{k+1}} \sum_{\nu = m^2+n^2} (m^2+n^2)^{1/3} | a_{mn} |^2,$$

which equals $\sum | a_{mn} |^2 (m + n^2)^{1/3}$, since the convergence is absolute. Thus $\sum_{k=0}^{\infty} \int\int | f(P) - S_{\nu_k} |^2 dx\, dy < \infty$; consequently $\sum_{k=0}^{\infty} | f(P) - S_{\nu_k} |^2 < \infty$ a. e. on C and $\lim_{k\to\infty} S_{\nu_k} = f(P)$ a. e. on C.

Proof of Theorem 2.2 [4]. From Lemma 2.3 it remains to prove that $\lim_{k\to\infty} (S_R(P) - S_{\nu_k}(P)) = 0$ a. e. for $\nu < R < \nu_{k+1}$, where from Lemma 2.3 $k \leq \nu_k^{1/3} < k+1$. By Lemma 2.2 there exists a monotone sequence $\{p_\nu\}$ with $p_\nu \to \infty$, such that $\sum p_\nu^2 \nu^{1/3} | a_{mn} |^2 < \infty$ $(\nu = m^2 + n^2)$. Setting $b_{mn} = p_\nu \nu^{1/6} a_{mn}$, $c_\nu = p_\nu \nu^{1/6}$, $\Phi_\mu = \sum_{m^2+n^2=\mu} b_{mn} e(m\alpha + n\beta)$ we have by Abel's transformation for $k > 0$

$$S_R - S_{\nu_k} = \sum_{\nu_k < \nu \leq R} \Big(\sum_{\mu=0}^{\nu} \Phi_\mu (c_\nu^{-1} - c_{\nu+1}^{-1}) - c_{\nu_k+1}^{-1} \sum_{0\leq\nu\leq\nu_k} \Phi_\nu + c_{R+1}^{-1} \sum_{0\leq\nu\leq R} \Phi_\nu.$$

Since $\sum | b_{mn} |^2 < \infty$, b_{mn} are the Fourier coefficients of a real Lebesgue square integrable function. Hence by Lemma 2.1 $\sum_{0\leq\nu\leq\nu_k} \Phi_\nu = O(\nu_k^{1/6})$ and therefore

$$| S_R - S_{\nu_k} | = O(R^{1/6}) (c_{\nu_k+1}^{-1} - c_{R+1}^{-1}) + O(p_{\nu_{k+1}}^{-1}) + O(p_{R+1}^{-1}).$$

Now $R\nu_k^{-1} \leq \nu_{k+1}\nu_k^{-1} \leq (k+2)^3 k^{-3} = O(1)$, so that $| S_R - S_{\nu_k} | = O(p_{\nu_k}^{-1})$, which $\to\infty$ as $k \to \infty$. Hence the theorem is proved.

3. The case $q = 3$. The result corresponding to Theorem 2.1 is

THEOREM 3.1. *The Lebesgue function* $L_R(\alpha, \beta, \gamma)$ *is* $O(R^{2/3}(\log R)^{5/3})$.

Proof. As in the proof of Theorem 2.1 we subdivide the cube $C(-\pi \leq x \leq \pi, -\pi \leq y \leq \pi, -\pi \leq z \leq \pi)$ by drawing planes parallel to the y- and z-axes through the points $-\pi + \epsilon$, $-\epsilon$, 0, ϵ, $\pi - \epsilon$, where ϵ is an arbitrary small positive number. Using the bound $K_R(x, y, z) = O(R^{3/2})$, we get that the integral of $| K_R(x, y, z) |$ over parallelepipeds of dimension $\epsilon \times \epsilon \times 2\pi$ is $O(\pi\epsilon^2 R^{3/2})$. For parallelepipeds of dimension $\epsilon \times \pi - 2\epsilon \times 2\pi$ analogously to (2.2), we use

$$(3.2) \quad K_R(x, y, z) = \frac{1}{e(z) - 1} \sum_{m^2+n^2 \leq R} e(mx + ny) \times \{ e(([(R - m^2 - n^2)^{\frac{1}{2}}] + 1)z) - e(-z[(R - m^2 - n^2)^{\frac{1}{2}}]) \} = O(R \csc \tfrac{1}{2} z),$$

if $| z | \geq \epsilon$, so that, for example,

$$\int_{-\pi}^{\pi} dx \int_{\epsilon}^{\pi} dz \int_{0}^{\epsilon} | K_R(x, y, z) | \, dy = O(\epsilon R \log \epsilon^{-1}).$$

Similarly as in (2.4),

$$(3.3) \quad \int_{-\pi}^{\pi} dx \int_{\epsilon}^{\pi-\epsilon} dz \int_{\alpha z - \frac{1}{2}\epsilon}^{\alpha z + \frac{1}{2}\alpha} | K_R(x, y, z) | \, dy = O(\epsilon R \log \epsilon^{-1}), \quad (\alpha \text{ a constant}).$$

In order to discuss the sum $\sum_{m^2+n^2+p^2\leq R}$ for parallelepipeds of dimension $\pi - 2\epsilon \times \pi - 2\epsilon \times 2\pi$, following Hua [3], we subdivide the sphere $m^2 + n^2 + p^2 \leq R$ by planes $m = 0$, $n = 0$, $p = 0$, $m = n$, $m = p$ and $n = p$. On the plane $n = p$, for example, we have

$$(3.4)\quad Q_1 = \sum_{m^2+2n^2\leq R} e(mx + (y+z)n)$$

$$= \frac{1}{e(y+z) - 1} \sum_{m=-[R^{1/2}]}^{[R^{1/2}]} e(mx)\{e(([(\tfrac{1}{2}(R - m^2))^{\frac{1}{2}}] + 1)(y+z)$$

$$- e(-[(\tfrac{1}{2}(R - m^2))^{\frac{1}{2}}](y+z))\}$$

$$= O(R^{\frac{1}{2}} \csc \tfrac{1}{2}(y+z)), \qquad (y + z \neq 0 \text{ or } \pm 2\pi),$$

so that as in (2.6) $\int_{-\pi}^{\pi} dx \int_{\epsilon}^{\pi-\epsilon} \int_{\epsilon}^{\pi-\epsilon} |Q_1|\, dy\, dz = O(R^{\frac{1}{2}} \log \epsilon^{-1})$. On the other hand over $(-\pi \leq x \leq \pi, \epsilon \leq y \leq \pi - \epsilon, -\pi + \epsilon \leq z \leq -\epsilon)$ we use the procedure of (3.3) with $\alpha = -1$ and (2.7) to obtain the same bound. Similar bounds are obtained on the planes $m = 0$, $n = 0$, $m = n$ and $m = p$.

The remaining sums to be considered in $\sum_{m^2+n^2+p^2\leq R}$ are all of the form

$$(3.5)\qquad \sum_{m^2+n^2+p^2\leq R} e(\pm mx \pm ny \pm pz)$$

$(-\pi \leq x \leq \pi, \epsilon \leq y \leq \pi - \epsilon, \epsilon \leq z \leq \pi - \epsilon) = P_\epsilon$, where $0 < m < n < p$ or this sum with m, n and p permuted. In considering (3.5) the discontinuous bound $O(|y - \alpha_m z|^{-1})$, where $\alpha_m = m(R - 2m^2)^{-\frac{1}{2}}$ $(0 < m \leq (R/3)^{-\frac{1}{2}})$, is encountered. This bound cannot be integrated over the domain

$$(3.6)\qquad D_{m\epsilon} = (-\pi \leq x \leq \pi, \epsilon \leq z \leq \pi - \epsilon, \alpha_m z - \tfrac{1}{2}\epsilon \leq y \leq \alpha_m z + \tfrac{1}{2}\epsilon),$$

but as for the case $q = 2$, where the bound $O(|y - z|^{-1})$ occurred, we replace $O(|y - \alpha_m z|^{-1})$ over $D_{m\epsilon}$ by $|K_R(x, y, z)|$ plus the sum of all other bounds used in considering (3.5) and notice that all of these bounds may be integrated over $D_{m\epsilon}$ to give the same bound or an even sharper result than the bound obtained over P_ϵ. Of course, since α_m depends on m it is necessary to integrate over $D_{m\epsilon}$ before summing with respect to m, which will be seen to be possible in all cases.

Defining $G(R)$ as the set $(0 < m \leq (R/3)^{\frac{1}{2}}, m < n < (\tfrac{1}{2}(R - m^2))^{\frac{1}{2}})$, we consider as in (2.9)

$$(3.7)\quad S = \sum_{(m,n)\varepsilon G(R)} \sum_{n<p<(R-m^2-n^2)^{1/2}} e(mx + ny + pz)$$

$$= \frac{e(z)}{e(z) - 1} \sum_{(m,n)\varepsilon G(R)} \{e(mx + ny + z[(R - m^2 - n^2)^{\frac{1}{2}}])$$

$$- e(mx + (y+z)n)\}$$

$= S_1 + S_2$. The sum S_2 may be treated the same way as in (3.4) and hence we get $\iiint\limits_{P_\epsilon} | S_2 | \, dx\, dy\, dz = O(R^{\frac{1}{2}} \log^2 \epsilon^{-1})$ (cf. (2.10)). We can use the same methods on the sum S_1 as were used on the corresponding sum for $q = 2$, that is, we consider

$$(3.8) \quad S_{11} = \frac{e(z)}{e(z) - 1} \sum_{(m,n) \varepsilon G(R)} e(mx + ny + z(R - m^2 - n^2)^{\frac{1}{2}} - \tfrac{1}{2}z)$$

and

$$(3.9) \quad S_{12} = \frac{e(z)}{e(z) - 1} \sum_{(m,n) \varepsilon G(R)} \{ e(mx + ny + z[(R - m^2 - n^2)^{\frac{1}{2}}]) \\ - e(mx + ny + z(R - m^2 - n^2)^{\frac{1}{2}} - \tfrac{1}{2}z) \}$$

where $S_1 = S_{11} + S_{12}$.

In S_{11} the function $2\pi f(u) = yu + z(R - m^2 - u^2)^{\frac{1}{2}}$ $(m < u < R_m$, $R_m = (\frac{1}{2}(R - m^2))^{\frac{1}{2}})$ satisfies the hypotheses of the van der Corput theorem for $(0 \leq y \leq \pi, 0 \leq z \leq \pi; 0 \leq m \leq R_0 = (R/3)^{\frac{1}{2}})$. Consequently

$$(3.10) \quad e(-\tfrac{1}{2}z)(e(z) - 1) S_{11} = \sum_{0 < m \leq R_0} \{ \int_m^{R_m} e(mx + uy \\ + z(R - m^2 - u^2)^{\frac{1}{2}}) \, du + O(1) \}. \\ = \sum_{0 < m \leq R_0} e(mx) I_m + O(R^{\frac{1}{2}}).$$

Using the same procedure as for $q = 2$, we show that

$$(3.11) \qquad \sum_{0 < m \leq R_0} \iiint\limits_{P_\epsilon - D_{m\epsilon}} \csc \tfrac{1}{2}z \, | I_m | \, dx\, dy\, dz = O(R^{\frac{1}{2}} \log^2 \epsilon^{-1}),$$

where $P_\epsilon - D_{m\epsilon}$ is the set of points belonging to P_ϵ but not to $D_{m\epsilon}$ and $\alpha_m = m(R - 2m^2)^{-\frac{1}{2}}$.

Proof. Setting $rv = uy + z(R - m^2 - u^2)^{\frac{1}{2}}$, where $r = (y^2 + z^2)^{\frac{1}{2}}$ and $ru = yv - z(R - m^2 - v^2)^{\frac{1}{2}}$, we get

$$I_m = \frac{1}{r} \int_{v_1}^{v_2} (y + \frac{zv}{(R - m^2 - v^2)^{\frac{1}{2}}}) e(rv) \, dv = I_{m1} + I_{m2},$$

where $v_1 = r^{-1}(my + z(R - 2m^2)^{\frac{1}{2}})$ and $v_2 = r^{-1}(y + z)R_m$. Now $I_{m1} = O(yr^{-2})$ and as in (2.15)

$$\sum_{0 < m \leq R_0} \iiint\limits_{P_\epsilon - D_{m\epsilon}} \csc \tfrac{1}{2}z \, | I_{m1} | \, dx\, dy\, dz \leq \sum_{0 < m \leq R_0} \iiint\limits_{P_\epsilon} \csc \tfrac{1}{2}z \, | I_{m1} | \, dx\, dy\, dz \\ = O(R^{\frac{1}{2}} \log^2 \epsilon^{-1}).$$

The function $g(v) = v(R - m^2 - v^2)^{-\frac{1}{2}}$ of I_{m2} is monotone and hence by the second mean value theorem

$$I_{m2} = zr^{-2}g(v_1)O(1) + zr^{-2}g(v_2)O(1),$$

where

$$g(v_1) = (my + z(R - 2m^2)^{\frac{1}{2}})(mz - y(R - 2m^2)^{\frac{1}{2}})^{-1}$$

and

$$g(v_2) = (y + z)\,|\,y - z\,|^{-1},$$

which equals $g(v_1)$ for $m = R_0$. As a function of y and z, $g(v_1)$ is discontinuous at $y = \alpha_m z$, where $R^{-\frac{1}{2}} < \alpha_m \leq 1$. Moreover for all such α_m the line $y = \alpha_m z$ crosses the square $(\epsilon \leq y \leq \pi - \epsilon, \epsilon \leq z \leq \pi - \epsilon)$ if ϵ is of the order $O(R^{-\frac{1}{2}})$ (cf. later). To avoid the discontinuity at $y = \alpha_m z$ we use (3.3) with $\alpha = \alpha_m$ and notice that all the other bounds are either $O(R^{\frac{1}{2}} \log^2 \epsilon^{-1})$ or $O(R^{5/6}\epsilon)$ (cf. (3.20)) over the domain $\sum_{0 < m \leq R_0} D_{m\epsilon}$. Over $P_\epsilon - D_{m\epsilon}$, evaluating as in (2.16) we get

$$\iiint_{P\epsilon - D_{m\epsilon}} zr^{-2} \csc \tfrac{1}{2}z(\alpha_m y + z)\,|\,\alpha_m z - y\,|^{-1}\,dx\,dy\,dz = O(\log^2 \epsilon^{-1}),$$

which proves (3.11) upon summing with respect to m.

As in the case $q = 2$ and since our proofs are independent of x, in addition to S of (3.7) it is sufficient to discuss that sum whose exponent is $mx + ny - pz$. In particular we need only consider

$$(3.12) \qquad T_1 = \frac{e(-\frac{1}{2}z)}{e(-z) - 1} \sum_{(m,n)\,\varepsilon\, G(R)} e(mx + ny - z(R - m^2 - n^2)^{\frac{1}{2}}), \qquad (x, y, z)\ \varepsilon\ P_\epsilon$$

(cf. (2.17)). As in the case $q = 2$ the van der Corput theorem fails for this sum but the method of proof is applicable. Hence by Euler's formula

$$\begin{aligned}(3.13) \quad T_1 = \frac{e(-\frac{1}{2}z)}{e(-z) - 1} \sum_{0 < m \leq R_0} \Big\{ &\int_m^{R_m} e(mx + uy - z(R - m^2 - u^2)^{\frac{1}{2}})\,du \\ &- e(mx + R_m(y - z))\,b(R_m) - \tfrac{1}{2}e(m(x + y) - z(R - m^2)^{\frac{1}{2}}) \\ &+ \int_m^{R_m} b(u)\,d(e(mx + uy - z(R - m^2 - u^2)^{\frac{1}{2}}))\Big\} \\ = \frac{e(-\frac{1}{2}z)}{e(-z) - 1}\,&(T_{11} + T_{12} + T_{13} + T_{14}).\end{aligned}$$

In this case the function $2\pi f(u) = yu - z(R - m^2 - u^2)^{\frac{1}{2}}$ has a bounded monotone derivative which does not change sign in (m, R_m); indeed

$$(2\pi)^{-1}\epsilon \leq f'(m) = (2\pi)^{-1}(y + \alpha_m z) \leq f'(u) \leq f'(R_m) = (2\pi)^{-1}(y + z),$$

so that $|f'(u)| \leq 1 - \epsilon(\pi)^{-1}$. Thus $1/f'(w)$ is bounded and monotone in the interval (m, R_m) and by the second mean value theorem

$$J_{1m} = \int_m^{R_m} (2\pi f'(u))^{-1} d(e(2\pi f(u))) = O(y^{-1}) + O((y+z)^{-1}),$$

since $f'(m) \geq (2\pi)^{-1} y$, so that

$$\underset{P\epsilon}{\int\int\int} \csc \tfrac{1}{2} z \,\Big|\sum_{0<m\leq R_0} e(mx) J_{1m}\Big|\, dx\, dy\, dz = O(R^{\frac{1}{2}} \log^2 \epsilon^{-1}).$$

Also, T_{12} and T_{13} in (3.13) are $O(R^{\frac{1}{2}})$. On T_{14} we use the same procedure as for (2.19) since the corresponding functions $g_1(u) = f'(u)(f'(u) + \nu)^{-1}$ and $g_2(u) = f'(u)(f'(u) - \nu)^{-1}$ are also monotone and bounded and do not change sign in (m, R_m). The only difference is that

$$g_2(m) = (y + \alpha_m z)(y + \alpha_m z - 2\pi\nu)^{-1},$$

but

$$|g_2(m)| \leq (y+z)(2\pi\nu - y - z)^{-1},$$

which is the value of $g_2(R_m)$ and similar to the case $q = 2$. Thus

$$\underset{P\epsilon}{\int\int\int} |T_1|\, dx\, dy\, dz = O(R^{\frac{1}{2}} \log^2 \epsilon^{-1}).$$

Now consider S_{12} of (3.9). Using the same argument as for the case $q = 2$,

$$\begin{aligned} e(z[(R - m^2 - n^2)^{\frac{1}{2}}]) &- e(z((R - m^2 - n^2)^{\frac{1}{2}} - \tfrac{1}{2})) \\ &= zb((R - m^2 - n^2)^{\frac{1}{2}}) O(1), \end{aligned}$$

so that

$$\begin{aligned} (3.14) \qquad S_{12} &= O\Big(z \csc \tfrac{1}{2} z \sum_{(m,n)\,\varepsilon\, G(R)} |b((R - m^2 - n^2)^{\frac{1}{2}})|\Big) \\ &= O\Big(z \csc \tfrac{1}{2} z \sum_{(m,n)\,\varepsilon\, G(R)} (\tfrac{1}{2} - \{(R - m^2 - n^2)^{\frac{1}{2}}\})\Big). \end{aligned}$$

But by a theorem due to Hua[3]

$$(3.19) \qquad \sum_{(m,n)\,\varepsilon\, G(R)} \{(R - m^2 - n^2)^{\frac{1}{2}}\} = \tfrac{1}{2} P + O(R^{2/3} (\log R)^{5/3}),$$

where $P = \sum_{(m,n)\,\varepsilon\, G(R)} 1$. Consequently

$$\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |S_{12}|\, dx\, dy\, dz = O(R^{2/3} (\log R)^{5/3}).$$

In connection with (3.11) we must show that

$$\sum_{(m,n)\,\varepsilon\, G(R)} \underset{D_{m\epsilon}}{\int\int\int} |b((R - m^2 - n^2)^{\frac{1}{2}})|\, dx\, dy\, dz$$

is sufficiently small. For fixed m the function $f(u) = 1 - (R - m^2 - n^2)^{\frac{1}{2}}$ satisfies the hypotheses of the theorem quoted on p. 8 where the z of the theorem may be taken to be $R^{1/6}$ again. Thus by subtraction we get

$$\sum_{m \leq n < R_m} \{(R - m^2 - n^2)^{\frac{1}{2}}\} = \tfrac{1}{2} \sum_{m \leq n < R_m} 1 + O(R^{1/3}),$$

or

$$\sum_{m < n < R_m} | b((R - m^2 - n^2)^{\frac{1}{2}}) | = O(R^{1/3}),$$

so that

$$\sum_{0 < m \leq R_0} \iiint_{D_{m\epsilon}} \sum_{m < n \leq R_m} | b((R - m^2 - n^2)^{\frac{1}{2}}) | \, dx\, dy\, dz = O(R^{5/6}\epsilon). \tag{3.20}$$

Since these proofs are independent of the signs in front of the x, y and z terms in $e(\pm mx \pm ny \pm pz)$, the proof of Theorem 3.1 is complete if we take $\epsilon = R^{-\frac{1}{2}}$. Since the sequence $\{(m^2 + n^2 + p^2)^{2/3}(\log(m^2 + n^2 + p^2))^{5/3}\}$ is monotonic increasing,, the proof of Theorem 2.2 generalizes without difficulty so that we have

THEOREM 3.2. *If the series*

$$\sum (m^2 + n^2 + p^2)^{2/3} (\log(m^2 + n^2 + p^2))^{5/3} \, | a_{mnp} |^2$$

is convergent, then $\lim_{R \to \infty} S_R$ *exists a. e. on* $(-\pi \leq x \leq \pi,\ -\pi \leq y \leq \pi,\ -\pi \leq z \leq \pi)$.

4. *The general case.* The following theorem is readily proved:

THEOREM 4.1. *The Lebesgue function* $L_R(\alpha_1, \cdots, \alpha_q)$ *is* $O(R^{\frac{1}{2}q})$.

Proof. The proof depends only on the boundedness and orthogonality properties of the exponential system. By the Schwarz inequality

$$\begin{aligned} L^2_R(\alpha_1, \cdots, \alpha_q) &= \{\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} | K_R(P - Q) | \, dA\}^2 \\ &= O(\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} | K_R(P - Q) |^2 \, dA), \end{aligned} \tag{4.1}$$

where $P = (\alpha_1, \cdots, \alpha_q)$, $Q = (x_1, \cdots, x_q)$ and $dA = dx_1 \cdots dx_q$. By the orthonormal properties (1.1)

$$\int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} | K_R(P - Q) |^2 \, dA = \sum_{\nu \leq R} 1, \quad \text{where } \nu = m^2_1 + \cdots + m^2_q.$$

But the right hand side of this equation is the number of lattice points in the hypersphere $m^2_1 + \cdots + m^2_q \leq R$, which is $O(R^{\frac{1}{2}q})$ [6]. Hence the conclusion of the theorem follows.

Using Theorem 4.1, it is easy to prove by the method of paragraph 2.4

THEOREM 4.2. *If* $\sum(m^2_1 + \cdots + m^2_q)^{\frac{1}{2}q} \mid a_{m_1 \ldots m_q} \mid^2 < \infty$, *then* $\lim_{R\to\infty} S_R$ *exists a. e. on* $(-\pi \leq x_k \leq \pi,\ k = 1, \cdots, q)$.

We remark that it would be impossible to improve this theorem by the methods used in Theorems 2.1 and 3.1 inasmuch as present results give the number of lattice points in the hypersphere

$$m^2_1 + \cdots + m^2_q \leq R \text{ as } \pi^{\frac{1}{2}q}(\Gamma(\tfrac{1}{2}q+1))^{-1}R^{\frac{1}{2}q} + O(R^{\frac{1}{2}q-1})$$

for $q \geq 5$ and as $\frac{1}{2}\pi^2 R^2 + O(R^{1+\epsilon})$ for $q = 4$ [7].

UNIVERSITY OF ILLINOIS.

REFERENCES.

[1] K. Chandrasekharan, "On the summation of multiple Fourier series III," *Bulletin of the American Mathematical Society*, vol. 52 (1946), pp. 474-477.

[2] ——— and S. Minakshisundaran, "Some results on double Fourier series," *Duke Mathematical Journal*, vol. 14 (1947), pp. 731-753.

[3] L. K. Hua and L. Schoenfeld, *Topics in Analytic Number Theory*, New York (to be published).

[4] S. Kaczmarz and H. Steinhaus, Theorie der Orthogonalreihen, *Monografie Matematyczne*, vol. VI, Warsaw (1935).

[5] K. Knopp, *Theory and Application of Infinite Series*, London and Glasgow (1928).

[6] E. Landau, *Vorlesungen über Zahlentheorie*, vol. 1 and 2, Leipzig (1927).

[7] A. Walfisz, "Über Gitterpunkte in mehrdimensionalen Ellipsoiden," *Mathematische Zeitschrift*, vol. 19 (1924), pp. 300-307.

[8] A. Zygmund, Trigonometrical Series, *Monografie Matematyczne*, vol. V, Warsaw (1935).

ORDER IN OPERATOR ALGEBRAS.*

By S. Sherman.

A partial order may be introduced among the hermitian elements of a C^* algebra $\mathcal{A}$ by letting $U \geqq 0$ mean that U is positive semi-definite. It is well known [7] [1, Lemma 1] that if a C^* algebra is commutative, the ordering (always the one given above) is a lattice ordering. We show here that if the hermitian elements of a self-adjoint algebra (with or without unit, no topological assumptions) of bounded operators on Hilbert space form a lattice, then the algebra is commutative. By example it is shown that lattice ordered subspaces of hermitian elements of $\mathcal{A}$ are not necessarily (commutative) algebras.

Let $\mathcal{A}$ be a C^* algebra [6, page 75], i. e. a uniformly closed self-adjoint complex algebra of bounded operators on a complex Hilbert space H. Thus if $U, V \varepsilon \mathcal{A}$ and α is a complex number, then UV, $U+V$, V^* (the adjoint of V) and αV are elements of $\mathcal{A}$. We define

$$\| U \| = \sup\{ (Ux, Ux)^{\frac{1}{2}} \mid (x, x) = 1, x \,\varepsilon\, H \}$$

and the requirement of uniform closure means that $\mathcal{A}$ is complete with respect to the norm $\| U \|$. Let $\mathcal{S} = \{U \mid U \varepsilon \mathcal{A}, U^* = U\}$ be the set of hermitian elements of $\mathcal{A}$. We say that $U \geqq 0$ if $(Ux, x) \geqq 0$ for each $x \,\varepsilon\, H$. This implies that $U \varepsilon \mathcal{S}$. Let $U > 0$ mean $U \geqq 0$ and $U \neq 0$.

We wish to investigate consequences of the assumption that $\mathcal{S}$ is a lattice, i. e. for every $U \varepsilon \mathcal{S}$, there is a $U_+ \geqq 0$ such that $U_+ - U \geqq 0$ and $U' \geqq U_+$ for any $U' \varepsilon \mathcal{S}$ that satisfies the two conditions $U' \geqq 0$ and $U' - U \geqq 0$. Denote by U_- the element $(-U)_+$ and by $| U |$ the element $U_+ + U_-$. Thus $U = U_+ - U_-$ and $| U | = U_+ + U_-$. In order not to obscure the essential lines of the proof with details we first handle the case with unit, i. e. the case where $\mathcal{A}$ contains the identity operator I.

Theorem 1. *If $\mathcal{S}$, the set of hermitian elements of $\mathcal{A}$, which is a C^* algebra with unit, is lattice ordered, then $\mathcal{A}$ is commutative.*

Proof. $\mathcal{S}$ satisfies Axioms I-V of [2], if the u, $\| U \|_u$, and E_u of [2] are I, $\| U \|$, and $\mathcal{S}$ respectively. A linear functional $\omega \mid \mathcal{S}$ is called positive $(\omega \geqq 0)$ if $U \geqq 0$ implies $\omega(U) \geqq 0$. As in [2] we define $\| \omega \| = \omega(I)$.

---

* Received December 1, 1949.

Let Ω be the set of all positive functionals having norm 1, i. e. $\omega \geqq 0$ and $\omega(I) = 1$. Let Π be the set of extreme points of Ω. Note: when the domain of $\omega \mid \mathcal{S} \in \Omega$ is extended in a natural way so that we get $\bar{\omega} \mid \mathfrak{A}$, Segal refers to $\bar{\omega}$ as a state and if $\omega \mid \mathcal{S} \in \Pi$, then $\bar{\omega} \mid \mathfrak{A}$ is called by Segal a pure state. We use this terminology. [2] refers to Ω and Π as H_u and S_u respectively. As in Theorem 3 of [2] we associate with each $U \in \mathcal{S}$ a function $\phi_U \in C(\Pi, R)$, where $\phi_U(\omega) = \omega(U)$ for each $\omega \in \Pi$ and $C(\Pi, R)$ is the collection of all continuous real-valued functions on the compact Hausdorff space Π (using w^* topology). According to [2] the one-one linear correspondence $U \leftrightarrow \phi_U$ possesses the following properties:

1) $\qquad U \leqq V \Longleftrightarrow \phi_U(\omega) \leqq \phi_V(\omega)$ for each $\omega \in \Pi$,

2) $\qquad \| U \| = \sup\{ | \phi_U(\omega) | \mid \omega \in \Pi \} \qquad (= df \, \| \phi_U \|)$,

3) $\qquad \phi_I(\omega) = 1$ for each $\omega \in \Pi$,

4) $\qquad \{ \phi_U \mid U \in \mathcal{S} \} = C(\Pi, R)$.

Let $\mathfrak{J}_0 = \{ U \mid \bar{\omega}_0(U) = 0 \}$, where $\bar{\omega}_0 \mid \mathcal{S} = \omega_0 \in \Pi$. We note that $\mathfrak{J}_0$ is a closed complex linear subspace of $\mathfrak{A}$ and we wish to show that $\mathfrak{J}_0$ is a two-sided ideal of $\mathfrak{A}$. For each $V \in \mathfrak{A}$, $V = V_R + iV_I$ where $V_R = (V + V^*)2^{-1}$, $V_I = (V - V^*)(2i)^{-1}$, $V_R \in \mathcal{S}$, and $V_I \in \mathcal{S}$. (We note in passing that if we let $\phi_V(\omega) = \phi_{V_R}(\omega) + i\phi_{V_I}(\omega)$, then $\{ \phi_V \mid V \in \mathfrak{A} \} = C(\Pi, C)$, the collection of all continuous complex-valued functions on Π.) In particular $U \in \mathfrak{J}_0 \Longrightarrow U_R \in \mathcal{S} \cap \mathfrak{J}_0$, $U_I \in \mathcal{S} \cap \mathfrak{J}_0$. Let $\mathcal{S}_+ = \{ V \mid \phi_V(\omega) \geqq 0$ for each $\omega \in \Pi \}$. For any $V \in \mathcal{S} \cap \mathfrak{J}_0$, $V = V_+ - V_-$, where $V_+ \in \mathcal{S}_+ \cap \mathfrak{J}_0$, $V_- \in \mathcal{S}_+ \cap \mathfrak{J}_0$, $\phi_{V_+}(\omega) = \sup\{\omega(V), 0\}$ and $\phi_{V_-}(\omega) = \sup\{-\omega(V), 0\}$. What we wish to show is that $U \in \mathfrak{J}_0$, $V \in \mathfrak{A} \Longrightarrow UV \in \mathfrak{J}_0$ and $VU \in \mathfrak{J}_0$. The above indicates that it is sufficient to show that $U \in \mathfrak{J}_0 \cap \mathcal{S}_+$, $V \in \mathfrak{A} \Longrightarrow UV \in \mathfrak{J}_0$ and $VU \in \mathfrak{J}_0$.

Suppose $U \in \mathfrak{J}_0 \cap \mathcal{S}_+$. Then by [1, page 201, Corollary 2] there exists a $W \in \mathcal{S}$ such that $W \geqq 0$ and $W^2 = U$. Obviously $\omega_0(VU) = \omega_0(ZW)$, where $Z = VW$, and by the Cauchy-Schwarz inequality

$$| \omega_0(ZW) | \leq (\omega_0(W^2))^{\frac{1}{2}} (\omega_0(ZZ^*))^{\frac{1}{2}} = 0.$$

From this it follows that $VU \in \mathfrak{J}_0$ and analogously $UV \in \mathfrak{J}_0$. Thus $\mathfrak{J}_0$ is a two-sided ideal in $\mathfrak{A}$. Since, for each $V \in \mathfrak{A}$, $\bar{\omega}_0(V - \bar{\omega}_0(V)I) = 0$ and so $V - \bar{\omega}_0(V)I \in \mathfrak{J}_0$, we see that $\bar{\omega}_0 \colon V \to \bar{\omega}_0(V)$ defines a homomorphism from $\mathfrak{A}$ to the complex numbers. Thus $\bar{\omega}_0(UV) = \bar{\omega}_0(U)\bar{\omega}_0(V) = \bar{\omega}_0(VU)$ and $\phi_{UV}(\omega_0) = \phi_{VU}(\omega_0)$ for each $\omega_0 \in \Pi$. Since $U \leftrightarrow \phi_U$ is one-one, $UV = VU$ for each U, $V \in \mathfrak{A}$, which proves our theorem. Incidentally since $\mathfrak{A}$ is commutative, $\mathcal{S}$ is a real commutative algebra. Note that the arguments would be valid also for a B^* algebra with unit [5, page 528], where with an appropriate definition of $\geqq$ we assume $U_1 \geqq U_2 \geqq 0 \Longrightarrow | U_1 | \geqq | U_2 | \geqq 0$.

The proof of the result analogous to Theorem 1 but without the assumption of a unit must contend with the possible distinction between the suprema (relative to imbedding ring) of a pair of elements $\{U_1, U_2\}$, imbedded in two rings one containing the other, $\{U_1, U_2\} \subset R_1 \subset R_2$. Furthermore the representation of $\mathcal{S}$ as a real linear space (closed under the operation of taking the absolute value) of real continuous functions on a compact Hausdorff space introduces some difficulty here. We abbreviate those portions of the proof which are duplications of the argument in Theorem 1.

LEMMA.

$$W_2 \geqq W_1 \geqq 0, \; 0 < \alpha_2 \leqq \alpha_1 \Longrightarrow (W_2)(\alpha_2 + W_2)^{-1} \geqq (W_1)(\alpha_1 + W_1)^{-1}.$$

Proof. From our hypothesis and the fact that all the expressions are only functions of W_2 we apply [1, page 201] and deduce that

$$(W_2)(\alpha_2 I + W_2)^{-1} \geqq (W_2)(\alpha_1 I + W_2)^{-1}.$$

We notice that $(z)(\alpha_1 + z)^{-1}$ is a "positive" [4, page 209] function, i. e. in the upper half of the complex plane the imaginary part of $(z)(\alpha_1 + z)^{-1}$ is non-negative. By [4, page 215], $(x)(\alpha_1 + x)^{-1}$ is a montone matrix function of arbitrarily high order in the interval $0 \leqq x < \infty$. By continuity considerations we deduce that $x(\alpha_1 + x)^{-1}$ is a monotone function for bounded positive semi-definite operators in Hilbert space and so $(W_2)(\alpha_1 + W_2)^{-1} \geqq (W_1)(\alpha_1 + W_1)^{-1}$. Our conclusion $(W_2)(\alpha_2 + W_2)^{-1} \geqq (W_1)(\alpha_1 + W_1)^{-1}$ now follows.[1]

THEOREM 2. *If $\mathcal{S}_0$, the set of hermitian elements of $\mathfrak{A}_0$, which is a C^* algebra without a unit, is lattice ordered, then $\mathfrak{A}_0$ is commutative.*

Proof. We consider the positive functionals $\omega \mid \mathcal{S}_0$, where now $\| \omega \| = \sup\{\omega(UU^*) \mid \| U \| \leqq 1\}$. As before Ω_0 is the set of positive functionals (states) having norm 1 and Π_0 is the set of extreme points (pure states) of Ω_0. We wish now to show that $\mathfrak{A}_0$ is isometric, linearly isomorphic and order isomorphic to a space $\mathfrak{C}(Q_0)$, i. e. a real linear space of real continuous functions on Q_0, the w^* closure of Π_0, which has the property that if ϕ is an element of $\mathfrak{C}(Q_0)$, then $|\phi|$, the function which for each $\omega \varepsilon Q_0$ takes the value $|\phi(\omega)|$, is an element of $\mathfrak{C}(Q_0)$.

---

[1] The referee has suggested the following abbreviated proof which is similar to one communicated to the author by Professor J. W. Givens: "This fact can readily be proved directly (without the use of Loewner's result or continuity considerations). It is readily seen to suffice to show that if A and B are bounded self-adjoint operators on a Hilbert space with bounded inverses and such that $A \geqq B > 0$, then $B^{-1} \geqq A^{-1}$. Now for any bounded operator C, $A \geqq B$ implies that $C^*AC \geqq C^*BC$, and taking C^{-1} to be a self-adjoint square root of B, we get $CAC \geqq I$, whence $C^{-1}A^{-1}C^{-1} \leqq I$, so $A^{-1} \leqq C^2 = B^{-1}$."

We readily verify that $\mathcal{A}_0$ satisfies Axioms I-IV of [3]. We now consider Axiom VI: $\| \, |U| \, \| = \| U \|$. Since $-|U| \leqq U \leqq |U|$, it follows that $\| U \| \leqq \| \, |U| \, \|$. From [1, page 201] we see that we can imbed U in $\mathcal{A}(U)$, the closed subring generated by U and I in $\mathcal{A}$, the C^* algebra obtained by adding a unit to $\mathcal{A}_0$. Since $\mathcal{A}(U)$ is a lattice we observe that $U^+ = \sup\{U, 0\}$ in $\mathcal{A}(U)$ and $U^- = \sup\{-U, 0\}$ in $\mathcal{A}(U)$ exist. Moreover $U^+ \varepsilon \mathcal{A}_0$, $U^- \varepsilon \mathcal{A}_0$, $U^+ \geqq U_+$, $U^- \geqq U_-$. Thus $U^+ + U^- \geqq U_+ + U_- = |U| \geqq 0$. But $\| U \| = \| U^+ + U^- \| \geqq \| \, |U| \, \| \geqq 0$, and so $\| U \| = \| \, |U| \, \|$, i. e., Axiom VI is satisfied.

In order to verify Axiom V:

$$U \geqq 0, \quad V \geqq 0 \Longrightarrow \| \sup\{U, V\} \| = \sup\{\| U \|, \| V \|\},$$

we define with [6, page 76], $\mu = \{U_1, \cdots, U_m\}$, $W(\mu) = \sum_{i=1}^{m} U_i U^*_i$, and $V(\mu) = W(\mu)(m^{-1}I + W(\mu))^{-1}$. The μ's form a system directed by set inclusion. From the Lemma, $\mu_\alpha \subset \mu_\beta \Longrightarrow V(\mu_\alpha) \leqq V(\mu_\beta) \leqq I$, where our elements are considered as elements of $\mathcal{A}$. Let

$$t(U;\mu) = \inf\{t \geqq 0 \mid -tV(\mu) \leq U \leqq tV(\mu)\},$$

where if no real t satisfying the inequality exists, we let $t(U;\mu) = +\infty$. If $\mu_\alpha \subset \mu_\beta$, then $t(U;\mu_\alpha) \geqq t(U;\mu_\beta) \geqq \| U \|$. If $U \geqq 0$, let $\mu = \{nU\}$, where n is a positive integer. Then

$$0 \leqq U \leqq (n^{-1} + \| U \|)(nU)(1 + nU)^{-1}.$$

Thus $\| U \| \leqq t(U;\mu) \leqq n^{-1} + \| U \|$ and $\lim_\mu t(U;\mu) = \| U \|$ for $U \geqq 0$. Now if $U \geqq 0$ and $V \geqq 0$ and $\epsilon > 0$ there exist $\mu_1(\epsilon)$ and $\mu_2(\epsilon)$ such that

$$0 \leqq U \leqq (\| U \| + \epsilon) V(\mu_1(\epsilon)), \qquad 0 \leqq V \leqq (\| V \| + \epsilon) V(\mu_2(\epsilon)).$$

If we let $\mu_3(\epsilon) = \mu_1(\epsilon) \cup \mu_2(\epsilon)$, then

$$0 \leqq \sup\{U, V\} \leqq (\sup\{\| U \|, \| V \|\} + \epsilon) V(\mu_3(\epsilon)).$$

From this $\| \sup\{U, V\} \| \leqq \sup\{\| U \|, \| V \|\}$. Since the reverse inequality is trivial we see that Axiom V is verified.

Now that Axioms I-VI of [3] are satisfied by $\mathcal{S}_0$, we deduce from [3, Theorem 3] and its proof that with each $U \varepsilon \mathcal{S}_0$ we may associate a $\phi_U \varepsilon C(Q_0, R)$ where Q_0, the w^* closure of Π_0, is compact and the mapping $U \rightarrow \phi_U$ is a one-one linear mapping into. Moreover

1′) $U \leq V \Longleftrightarrow \phi_U(\omega) \leqq \phi_V(\omega)$ for each $\omega \in Q_0 \Longleftrightarrow \phi_U(\omega) \leqq \phi_V(\omega)$ for each $\omega \in \Pi_0$,

2′) $\| U \| = \sup\{ | \phi_U(\omega) | \mid \omega \in Q_0\} = \sup\{ | \phi_U(\omega) | \mid \omega \in \Pi_0\}$,

3′) $\sup\{\phi_{UU^*}(\omega) \mid \| U \| \leq 1\} = 1$,

4′) $\phi \in \{\phi_U \mid U \in \mathcal{S}_0\} \Longrightarrow | \phi | \in \{\phi_U \mid U \in \mathcal{S}_0\}$,

where $| \phi | (\omega) = | \phi(\omega) |$ for each $\omega \in \Pi_0$ $(\omega \in Q_0)$.

Property 4′) is important since it implies that

$$\phi \in \{\phi_U \mid U \in \mathcal{S}_0\} \Longrightarrow (\phi + | \phi |)2^{-1} \in \{\phi_U \mid U \in \mathcal{S}_0\}.$$

Note that $\phi_{U_+}(\omega) = ((\phi + | \phi |)2^{-1})(\omega)$.

For each $\omega \in \Pi_0$, let $\mathfrak{J}_0 = \{U \mid U \in \mathcal{A}_0, \overline{\omega}_0(U) = 0\}$. Just as before $\mathfrak{J}_0$ is a closed two sided ideal (now in $\mathcal{A}_0$). A crucial fact in the proof is that $U \in \mathfrak{J}_0 \Longrightarrow U = U_+ - U_-$ where $U_+ \geqq 0$, $U_- \geqq 0$, $U_+ \in \mathfrak{J}_0$, and $U_- \in \mathfrak{J}_0$. We take care of this by using the characterization of ϕ_{U_+} given in the last paragraph. $\mathfrak{J}_0$ is also a two-sided ideal in $\mathcal{A}$ ($\mathcal{A}_0$ augmented in standard fashion with I). As before we deduce that $\overline{\omega}_0 \colon V \rightarrow \overline{\omega}_0(V)$ defines a homomorphism from $\mathcal{A}$ to the complex numbers and that for each $\omega_0 \in \Pi_0$, $\overline{\omega}_0(UV) = \overline{\omega}_0(U)\, \overline{\omega}_0(V) = \overline{\omega}_0(VU)$. Since Π_0 is dense in Q_0, we infer that $\phi_{UV} = \phi_{VU}$ and $UV = VU$. Thus Theorem 2 is proved.

If we examine the proofs of Theorems 1 and 2 we notice that we could dispense with completeness. The relevant representation theorems without completeness hypotheses can be found in [2, 3]. The intermediate constructions we make using completeness cause no essential difficulty.

COROLLARY. If $\mathcal{A}$ is a self-adjoint algebra (with or without unit, no completeness assumption) of bounded operators on a complex Hilbert space and if $\mathcal{S}$, the set of hermitian elements of $\mathcal{A}$, is a lattice under our ordering, then $\mathcal{A}$ is commutative.

Let $\mathcal{A}^n$ be the C^* algebra of all linear operators on H^n, n-dimensional complex euclidean space, and let $\mathcal{S}^n$ be the set of hermitian elements of $\mathcal{A}^n$. We are concerned with the possibility that if $\mathcal{S}$, a real linear subspace of $\mathcal{S}^n$, were a lattice under its natural ordering, then $\mathcal{S}$ would have to be a real commutative algebra. The facts are as follows. When $n = 2$ and $\mathcal{S}$ does not contain a unit, then $\mathcal{S}$ need not be a commutative algebra. For instance, choose a coordinate system and let $\mathcal{S}$ be the real linear subspace formed by the matrices $\{A_\alpha \mid -\infty < \alpha < \infty\}$, where

$$A_\alpha = \begin{pmatrix} \alpha & 0 \\ 0 & 2\alpha \end{pmatrix}.$$

When $n = 2$ and $\mathcal{S}$ contains the unit then $\mathcal{S}$ is a real commutative algebra. When $n \geqq 3$, $\mathcal{S}$ need not be a real commutative algebra even though $\mathcal{S}$ contains the identity operator. For example let $n = 3$ and after choosing a coordinate system let $\mathcal{S}$ be the real linear space represented by matrices of the form $\{\alpha A + \beta B \mid -\infty < \alpha, \beta < \infty\}$, where

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$\mathcal{S}$ is not an algebra since $AB \notin \mathcal{S}$.

It is readily checked that $\alpha A + \beta B$ is positive semi-definite if and only if $\alpha \geqq 0$, $\beta \geqq 0$. This makes $\mathcal{S}$ a well-known vector lattice.

Remark. Conversations with Professor Lipman Bers on the proof of the known fact that the collection of all real symmetric 2×2 matrices is not a lattice, led to the technique for constructing the counter-examples above and also raised the questions which they answer.

THE INSTITUTE FOR ADVANCED STUDY,
PRINCETON, NEW JERSEY.

BIBLIOGRAPHY.

[1] I. Gelfand and M. Neumark, "On the imbedding of normed rings into the ring of operators in Hilbert space," *Recueil Mathématique* (*Mat. Sbornik*), New Series, vol. 12 (1943), pp. 197-213.

[2] M. Krein and S. Krein, "On an inner characteristic of the set of all continuous functions defined on a bicompact Hausdorff space," *Comptes Rendus* (*Doklady*) *de l'Académie des Sciences de l'URSS*, New Series, vol. 27 (1940), pp. 427-30.

[3] ———, "Sur l'espace des fonctions continues définies sur un bicompact de Hausdorff et ses sous-espaces semiordonnés," *Recueil Mathématique* (*Mat. Sbornik*), New Series, vol. 13 (1943), pp. 1-38.

[4] K. Loewner, "Über monotone Matrixfunktionen," *Mathematische Zeitschrift*, vol. 38 (1934), pp. 177-216.

[5] C. E. Rickart, "Banach algebras with an adjoint operation," *Annals of Mathematics*, vol. 47 (1946), pp. 528-550.

[6] I. E. Segal, "Irreducible representations of operator algebras," *Bulletin of the American Mathematical Society*, vol. 53 (1947), pp. 73-88.

[7] M. H. Stone, "A general theory of spectra I," *Proceedings of the National Academy of Sciences*, vol. 26 (1940), pp. 280-3.

CALCUTTA UNIVERSITY CENTRAL LIBRARY.

THE THEORY OF UNITS OF QUADRATIC AND HERMITIAN FORMS.*

By K. G. RAMANATHAN.

1. Introduction.

1. Introduction. In a recent paper Humbert [7][1] carried through Siegel's reduction theory to indefinite quadratic or hermitian forms with coefficients which are integers of an algebraic number field K of finite degree over the field of rationals. He proved, among other theorems, the theorem that the number of classes of forms of given determinant, of order m, and with integral coefficients in K is finite.

It is the object of this paper to study the group of units of quadratic (hermitian) forms with integral elements in a finite algebraic extension K of the rationals. A unit of a form is a unimodular transformation leaving invariant the form. We prove

THEOREM A. *Let K be a totally real algebraic number field of finite degree over the field of rationals and $\mathfrak{S}$ the matrix of a quadratic form with elements integral in K. Let Γ be the group of units $\mathfrak{U}$ such that $\mathfrak{U}'\mathfrak{S}\mathfrak{U} = \mathfrak{S}$. Then* 1) *$\Gamma$ is finite if and only if $\mathfrak{S}$ is a totally definite symmetric matrix,* 2) *Γ is always finitely generated.*

The first part of Theorem A is proved in Section **4** and the second part in Section **5** since its proof requires the properties of fundamental domains. A similar theorem is proved in Section **4** for hermitian forms. It is shown that the theory of hermitian forms works only in fields of the type $K_0 = K(d^{1/2})$ where K is a totally real field and d is a totally negative number of K.

We then generalise Siegel's results [8] on the measure of unit groups. We obtain for Γ/C where C is the centre of Γ, a faithful and discontinuous representation as a group B of non-Euclidean motions leaving invariant an open subspace J of a space of M dimensions, M being an integer depending on the signature of the quadratic form. In J we define a volume element dv invariant under B. Since B is discontonuous in J we can construct a fundamental domain F which is shown to have a finite number of neighbours.

---

* Received September 9, 1949; revised and substantially enlarged, March 30, 1950; slightly modified, October 12, 1950.

[1] Numbers in square brackets refer to bibliography at the end.

The important theorem stated as Theorem B asserts that $\int_F dv$ is finite except when $\mathfrak{S}$ is the matrix of a decomposable binary quadratic form. In this case it is shown that $\int_F dv$ is infinite. The proof of this depends on certain ideas used by Blumenthal [2] in the construction of a fundamental domain for the Hilbert modular group.

It is possible to prove an analogue of Theorem B for quadratic forms over number fields which are not necessarily totally real. These and other considerations regarding representations by indefinite quadratic forms will be shown elsewhere.

Finally I should like to record my thanks to Professor C. L. Siegel for his kindness in reading through the original version of the paper and making useful suggestions.

2. Notations. Symmetric or hermitian matrices will be denoted by capital German letters $\mathfrak{S}, \mathfrak{H} \cdots$ etc. $\mathfrak{U}$ will denote unimodular matrices, that is matrices whose elements are integers of K and whose determinant is an algebraic unit of K. $|\mathfrak{A}|$ will denote the determinant of a matrix $\mathfrak{A}$. abs $\mathfrak{A}$ will denote the absolute value of the determinant of $\mathfrak{A}$. We shall use abs a also to denote the absolute value of a number a. Small German letters $\mathfrak{x}, \cdots$ will be used as column vectors. $\mathfrak{S}[\mathfrak{U}]$ and $\mathfrak{S}\{\mathfrak{U}\}$ stand for $\mathfrak{U}'\mathfrak{S}\mathfrak{U}$ and $\mathfrak{U}'\mathfrak{S}\bar{\mathfrak{U}}$ where $\mathfrak{U}'$ and $\bar{\mathfrak{U}}$ are transposed and complex conjugate of $\mathfrak{U}$. $N(\mathfrak{T})$ denotes the product of $\mathfrak{T}$ and all its conjugates under the isomorphisms of K. $N(\mathfrak{T})$ will be the usual norm if $\mathfrak{T}$ is composed of elements of K. Similarly $\sigma(\mathfrak{T})$ is the spur of $\mathfrak{T}$. If $\mathfrak{S}$ is a matrix, we mean by the system $(\mathfrak{S})$, a matrix vector $(\mathfrak{S}^{(1)}, \cdots, \mathfrak{S}^{(f)})$. $\mathfrak{E}_k$ will denote unit matrix of order k. $\mathfrak{O}$ will denote a zero matrix.

3. Reduction theory. Let K be an algebraic number field of finite degree $n = n_1 + 2n_2$ over the field of rationals and let $K^{(1)}, \cdots, K^{(n_1)}$ be the real conjugates of K and $K^{(n_1+\nu)}$, $K^{(n_1+n_2+\nu)}$, $\nu = 1, \cdots, n_2$ the pairs of complex conjugates of K. The unimodular matrices of order m in K form a multiplicative group Ω whose centre C consists of the matrices $\omega\mathfrak{E}_m$ where ω is a root of unity in K. Let Γ be the factor group Ω/C. The problem of reduction consists in finding for Γ a fundamental domain in a suitable space. This problem was solved by Humbert [6] in the following way.

Let $\mathfrak{S}^{(1)}, \cdots, \mathfrak{S}^{(n_1)}$ be the matrices of n_1 non-singular positive definite quadratic forms in m variables and $\mathfrak{S}^{(n_1+\nu)}$, $\mathfrak{S}^{(n_1+n_2+\nu)}$; $(\nu = 1, \cdots, n_2)$ pairs of complex conjugate positive definite hermitian non-singular matrices of order m. Let $\mathfrak{S}^{(i)}$, $(i = 1, \cdots, n_1 + n_2)$, be attached to the field $K^{(i)}$ and consider the system $(\mathfrak{S}) = (\mathfrak{S}^{(1)}, \cdots, \mathfrak{S}^{(n_1+n_2)})$. This system $(\mathfrak{S})$ defines a point in an open subspace P of a space of $\frac{1}{2}n_1 m(m+1) + n_2 m^2$ real dimensions. If $\mathfrak{U} \varepsilon \Gamma$, the mapping $(\mathfrak{S}) \rightarrow (\mathfrak{U}'\mathfrak{S}\bar{\mathfrak{U}})$ leaves the space P invariant. Since the algebraic units of K form a discrete subgroup of the continuous group in the space of $n_1 + n_2$ dimensions it follows that the representation $(\mathfrak{S}) \rightarrow (\mathfrak{S}\{\mathfrak{U}\})$ is discontinuous in P. Therefore we can construct a fundamental domain R in P for this discrete group of mappings. Let $\mathfrak{A}$ be a non-singular integral matrix in K and let $(\mathfrak{S}^*) = (\mathfrak{S}\{\mathfrak{A}\})$ be a system satisfying certain minimum conditions of Humbert. Then $\mathfrak{A}$ belongs to a finite set of matrices $\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_\mu$ which are uniquely determined, up to multiplication on the left by a unimodular matrix $\mathfrak{U}$, by the field K and the integer m. For any system there is, in general, associated only one $\mathfrak{A}_i$. Since $\mathfrak{A} = \mathfrak{U}\mathfrak{A}_i$ for some $\mathfrak{A}_i$ and a unimodular $\mathfrak{U}$, we consider the system $(\mathfrak{S}^{**}) = \mathfrak{S}^*\{\mathfrak{A}_\nu^{-1}\}$. The totality of these points $(\mathfrak{S}^{**})$ constitute in P a *fundamental domain* R. The points $(\mathfrak{S}^*)$ define a point set R_0 called the *Humbert domain* defined in (35), (36), (37). Humbert showed that R is a union of μ convex pyramids whose faces have equations of the form $\tau(\mathfrak{S}\{\mathfrak{x}\}) = \tau(\mathfrak{S}\{\mathfrak{y}\})$, where $\mathfrak{x}$ and $\mathfrak{y}$ are independent column vectors in K. Putting $\mathfrak{x} = \mathfrak{a} + \mathfrak{b}$ and $\mathfrak{y} = \mathfrak{a} - \mathfrak{b}$ we write the above equation as

$$\tau(\mathfrak{a}'\mathfrak{S}\bar{\mathfrak{b}} + \bar{\mathfrak{a}}'\mathfrak{S}\mathfrak{b}) = 0, \tag{1}$$

where $\mathfrak{a} \neq \alpha\mathfrak{b}$ and α is a pure imaginary quantity.

In a recent paper [7] Humbert carried through the reduction theory to indefinite quadratic forms over algebraic number fields. Let us consider quadratic forms over totally real fields.[2] Let K be a totally real algebraic number field of finite degree f over the rationals and $K^{(1)} = K, \cdots, K^{(f)}$ its conjugates. Let $\mathfrak{S}^{(1)} = \mathfrak{S}$ be the matrix of a quadratic form in m variables and with integral elements in K. Let n_r, m_r $(= m - n_r)$ be the signature of $\mathfrak{S}^{(r)}$ where $\mathfrak{S}^{(r)}$ is the image of $\mathfrak{S}^{(1)}$ under the isomorphism $K^{(1)} \rightarrow K^{(r)}$, $(r = 1, \cdots, f)$. For each r there exists a real matrix $\mathfrak{C}^{(r)}$ so that

$$\mathfrak{S}^{(r)} = \mathfrak{C}^{(r)\prime} \begin{pmatrix} \mathcal{E}_{n_r} & \mathfrak{O} \\ \mathfrak{O} & -\mathcal{E}_{m_r} \end{pmatrix} \mathfrak{C}^{(r)}.$$

Following Hermite we associate with $\mathfrak{S}^{(r)}$ a positive definite matrix $\mathfrak{H}^{(r)} = \mathfrak{C}^{(r)\prime}\mathfrak{C}^{(r)}$ so that $|\mathfrak{H}^{(r)}| = \operatorname{abs} \mathfrak{S}^{(r)}$. Also $\mathfrak{H}^{(r)}$ satisfies

$$\mathfrak{H}^{(r)}\mathfrak{S}^{(r)-1}\mathfrak{H}^{(r)} = \mathfrak{S}^{(r)}, \qquad (r = 1, \cdots, f). \tag{2}$$

---

[2] The case of arbitrary number fields will be dealt with elsewhere.

We thus obtain a positive system $(\mathfrak{H})$ attached to $(\mathfrak{S})$. We consider all the positive systems $(\mathfrak{H})$ in the space P which satisfy the equations (2). They define a subspace H which is an irreducible algebraic manifold of

$$M = \sum_{r=1}^{f} n_r m_r \tag{3}$$

real dimensions. The points of H can be expressed in terms of M real parameters as follows.

$$\mathfrak{H}^{(r)} = 2\mathfrak{W}^{(r)} - \mathfrak{S}^{(r)}; \qquad \mathfrak{W}^{(r)} = \mathfrak{T}^{(r)-1}[\mathcal{E}\ \mathfrak{Y}^{(r)\prime}];$$
$$\mathfrak{T}^{(r)} = \mathfrak{S}^{(r)-1}\begin{bmatrix}\mathcal{E}\\ \mathfrak{Y}^{(r)}\end{bmatrix} > 0, \qquad (r = 1, \cdots, f), \tag{4}$$

where $\mathfrak{Y}^{(r)}$ is a real matrix of n_r columns and m_r rows. (4) shows that the space

$$J\colon\ \mathfrak{T}^{(r)} > 0; \qquad (r = 1, \cdots, f) \tag{5}$$

of M real dimensions is mapped *upon* the space H of positive systems $(\mathfrak{H})$ satisfying (2). Let $\mathfrak{U}$ be a unimodular matrix in K such that $(\mathfrak{H}[\mathfrak{U}])$ is a point in the reduced space R; then we call $(\mathfrak{S}[\mathfrak{U}])$ *reduced*. Humbert showed [7] that for given value of $N(\mathfrak{S})$ the number of reduced integral systems $(\mathfrak{S})$ is finite.

In the case of hermitian forms a similar result can be proved. Let K_0 be an algebraic number field and $\mathfrak{N}$ its splitting field [1]. Then $\mathfrak{N}$ is a normal extension of the rationals. Let σ be an isomorphism of K_0 into its conjugate field K_1; then σ can be extended into an automorphism of $\mathfrak{N}$. Let τ be an automorphism of $\mathfrak{N}$ taking any quantity into its complex conjugate. A hermitian matrix with coefficients in K_0 is characterized by the property

$$a_{ij} = \bar{a}_{ji}, \tag{6}$$

so that for a hermitian matrix to go into a hermitian matrix by the isomorphism taking K_0 into its conjugate, it is necessary and sufficient that

$$\sigma\tau = \tau\sigma \tag{7}$$

for all automorphisms σ of $\mathfrak{N}$. Since τ is an element of order 2 we see that there is a subfield $\mathfrak{N}_0$ of $\mathfrak{N}$ which is invariant elementwise under $(1; \tau)$ where 1 is the identity automorphism. $\mathfrak{N}_0$ is a real subfield of $\mathfrak{N}$. If K_0 is real then by (7) all conjugates of K_0 would be real and our hermitian form would be merely a real quadratic form. Let us therefore take K_0 complex; then by (7) all its conjugates are complex. Let $\mathfrak{G} = K_0 \cap \mathfrak{N}$. Then $\mathfrak{G}$

is a real subfield of $\mathfrak{R}$. Also K_0 is of degree 2 over $\mathfrak{G}$. Let $\sigma_i K_0 = K_i$ for some automorphism σ_i; then K_i is of degree 2 over $\sigma_i \mathfrak{G}$ which is real. Thus $\mathfrak{G}$ is a totally real field and K_0 is an imaginary quadratic extension of $\mathfrak{G}$ such that all the conjugates of K_0 are complex. Thus $K_0 = K(d^{1/2})$ where K is totally real and d is a totally negative number of K. *It is therefore only in such fields that the theory of hermitian forms works.*

Let therefore K be a totally real field of degree $f \geqq 1$ over the rationals and d a totally negative integer of K. Let $K^{(1)} (= K), \cdots, K^{(f)}$ be the conjugates of K. Let $K_0{}^{(r)} = K^{(r)}(d_r{}^{1/2})$ and $K_0{}^{(r+f)} = K^{(r)}(-d_r{}^{1/2})$ so that $\bar{K}_0{}^{(r)} = K_0{}^{(r+f)}$. Let $\mathfrak{S}^{(r)}$ be a hermitian matrix with integral elements in K_0 and of signature n_r, m_r. As before we find that the system $(\mathfrak{S}) = (\mathfrak{S}^{(1)}, \cdots, \mathfrak{S}^{(f)})$ has attached with it the totally positive system $(\mathfrak{H}) = (\mathfrak{H}^{(1)}, \cdots, \mathfrak{H}^{(f)})$, where

$$\mathfrak{H}^{(r)} \mathfrak{S}^{(r)-1} \mathfrak{H}^{(r)} = \mathfrak{S}^{(r)}, \qquad (r = 1, \cdots, f)). \tag{8}$$

These systems $(\mathfrak{H})$ constitute an algebraic manifold H_1 of $2M$ real dimensions and the points of H_1 can be parametrized by means of M complex quantities ($2M$ real quantities) $\mathfrak{Z}^{(1)}, \cdots, \mathfrak{Z}^{(f)}$ satisfying

$$\mathfrak{T}^{(r)} = \mathfrak{S}^{(r)-1} \left\{ \begin{matrix} \mathcal{E} \\ \mathfrak{Z}^{(r)} \end{matrix} \right\} > 0, \qquad (r = 1, \cdots, f). \tag{9}$$

4. Group of units. Let K be a totally real field and $\mathfrak{S}$ the matrix of a quadratic form with integral coefficients. Let $\mathfrak{U}$ be a unimodular matrix such that

$$\mathfrak{U}' \mathfrak{S} \mathfrak{U} = \mathfrak{S}. \tag{10}$$

This implies that $\mathfrak{U}^{(r)\prime} \mathfrak{S}^{(r)} \mathfrak{U}^{(r)} = \mathfrak{S}^{(r)}$. Obviously $|\mathfrak{U}| = \pm 1$. We call $\mathfrak{U}$ *a unit of* $\mathfrak{S}$. There will be no confusion if we speak of the system $(\mathfrak{U}) = (\mathfrak{U}^{(1)}, \cdots, \mathfrak{U}^{(f)})$ as a unit of the system $(\mathfrak{S})$. The units $\mathfrak{U}$ obviously constitute a multiplicative group Γ_1 called the *group of units of* $\mathfrak{S}$.

Let us call a quadratic form $\mathfrak{x}' \mathfrak{S} \mathfrak{x}$ with integral elements in K totally definite if it and all its conjugates are definite forms. We shall now prove the

THEOREM 1. *Γ_1 is finite if and only if $\mathfrak{S}$ is the matrix of a totally definite form.*

To prove this theorem we require certain auxiliary lemmas.

LEMMA 1.[3] *A totally definite form* $\mathfrak{x}'\mathfrak{S}\mathfrak{x} = \sum_{i=1}^{m} a_i x_i^2$, *where* a_i *are integers of* K *has only a finite number of units.*

Proof. Let $\mathfrak{U}$ be a unimodular matrix so that $\mathfrak{U}'\mathfrak{S}\mathfrak{U} = \mathfrak{S}$. Let $\begin{pmatrix} u_{1k} \\ \vdots \\ u_{mk} \end{pmatrix}$ be the k-th column of $\mathfrak{U}$. Then since $\mathfrak{U}$ is a unit

$$\sum_{r=1}^{m} a_r u_{rk}^2 = a_k, \qquad (k = 1, \cdots, m). \tag{11}$$

Since the quadratic form is totally definite, a_r/a_s for all r and s are totally positive numbers of K. Thus

$$u_{rk}^2 \leq a_k/a_r \leq \sigma[(a_1 + \cdots + a_m)(a_1^{-1} + \cdots + a_m^{-1})]. \tag{12}$$

Thus we get

$$\operatorname{abs} u_{rk}^{(s)} \leq \{\sigma[(a_1 + \cdots + a_m)(a_1^{-1} + \cdots + a_m^{-1})]\}^{1/2} \tag{13}$$

for $s = 1, \cdots, f$. Thus by a well-known theorem of Kronecker the u_{rk} and hence $\mathfrak{U}$ have only a finite number of possibilities.

The Lemma follows at once if we observe that the group of all real linear transformations leaving a totally definite system $(\mathfrak{S})$ invariant is a compact Lie group and hence any discrete subgroup of it, like the group of units, must be finite.

LEMMA 2. *If a and b are integers of K and ab is not the square of an integer of K then the equation*

$$x^2 - aby^2 = 1 \tag{14}$$

has a finite or infinite number of solutions x, y integral in K according as ab is or is not totally negative.

This is a generalisation of the ordinary Pellian equation to algebraic number fields. The result is due to R. Fricke [4].

LEMMA 3. *Let a and b be integers of K and ab not a square. Then* $ax^2 - by^2$ *has an infinity of units unless* $(-ab)$ *is totally positive.*

Proof. If $-ab$ is totally positive then $ax^2 - by^2$ is totally definite and by Lemma 1 it has only a finite number of units. Let therefore $-ab$ be not totally positive. We now follow the method of Dirichlet-Dedekind [3].

---

[3] Here and elsewhere we write K for $K^{(1)}$ and $\mathfrak{S}$ for $\mathfrak{S}^{(1)}$. We use K in the theory of quadratic forms and K_0 in the theory of hermitian forms. When no superscripts appear, (11) etc. imply the corresponding relations got by taking the images in conjugate fields.

Let $x \to \alpha x + \beta y$; $y \to \gamma x + \delta y$, $\alpha, \beta, \gamma, \delta$ integers of K, be a linear transformation leaving invariant the form $ax^2 - by^2$. Then

$$(15) \qquad a\alpha^2 - b\gamma^2 = a; \qquad a\beta^2 - b\delta^2 = -b; \qquad a\alpha\beta - b\gamma\delta = 0.$$

From (15) we deduce that

$$(15') \qquad \alpha = \delta = p, \qquad \gamma = aq, \qquad \beta = bq,$$

where p, q are integers of K satisfying

$$(16) \qquad p^2 - abq^2 = 1,$$

satisfy (15).[4] If ab is not a square in K, then from Lemma 2, there is an infinity of integral solutions of (16). Thus the number of solutions of (15) is infinite.

We can now prove Theorem 1. Let $\mathfrak{x}'\mathfrak{S}\mathfrak{x}$ be a non-degenerate quadratic form with integral elements in K. There exists a matrix $\mathfrak{C}$ in K (not necessarily integral) such that $\mathfrak{C}'\mathfrak{S}\mathfrak{C} = \mathfrak{D}$ is a diagonal matrix. If $\mathfrak{U}$ is a unit of $\mathfrak{S}$ then $\mathfrak{C}^{-1}\mathfrak{U}\mathfrak{C} = \mathfrak{V}$ satisfies $\mathfrak{V}'\mathfrak{D}\mathfrak{V} = \mathfrak{D}$. But $\mathfrak{V}$ is not in general integral. Let q be a rational integer such that $q\mathfrak{C}$ and $q\mathfrak{C}^{-1}$ are integral matrices. The matrices $\mathfrak{V} \equiv \mathcal{E} \pmod{q^2}$ form a subgroup of finite index in $\mathfrak{C}^{-1}\Gamma_1\mathfrak{C}$. The integral matrices $\mathfrak{V} \equiv \mathcal{E} \pmod{q^2}$ form a subgroup of finite index in the group of all units $\mathfrak{V}$. Thus the finiteness or otherwise of Γ_1 can be established from that of the units of the quadratic form $\mathfrak{x}'\mathfrak{D}\mathfrak{x}$. We may, as is readily seen, assume that the elements of $\mathfrak{D}$ are integers of K. Let therefore $\mathfrak{x}'\mathfrak{D}\mathfrak{x} = a_1x_1{}^2 + \cdots + a_mx_m{}^2$, $a_1, \cdots, a_m$ being integers of K. If $\mathfrak{x}'\mathfrak{D}\mathfrak{x}$ is totally definite then Lemma 1 shows that Γ_1 is finite. Let $\mathfrak{x}'\mathfrak{D}\mathfrak{x}$ be not totally definite. Then at least 2 of the quantities, say a_1 and a_2, are such that a_1a_2 is not totally positive. Let $\mathfrak{W}$ be a unimodular matrix of order $m-2$ of a linear transformation leaving $a_3x_3{}^2 + \cdots + a_mx_m{}^2$ invariant. The set of $\mathfrak{W}$ is not null since $\mathfrak{W} = \mathcal{E}$ is a possible value of $\mathfrak{W}$. Let $\mathfrak{V}$ be a unit of $a_1x_1{}^2 + a_2x_2{}^2$. Then

$$\mathfrak{U} = \begin{pmatrix} \mathfrak{V} & \mathfrak{O} \\ \mathfrak{O} & \mathfrak{W} \end{pmatrix}$$

is a unit of $\mathfrak{x}'\mathfrak{D}\mathfrak{x}$. Lemma 3 shows that if $-a_1a_2$ is not the square of an integer of K, there are an infinity of $\mathfrak{V}$. Suppose now that—a_1a_2 is the square of an integer of K. Then as before it suffices to show that $2x_1x_2$ has an infinity of units. Indeed if ϵ is an algebraic unit of K then

---

[4] The solutions (15′) are not all the solutions of (15). What we are concerned with is the finiteness or otherwise of the number of integral solutions and for this the crude results (15′) are enough.

$$\mathfrak{B}_1 = \begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon^{-1} \end{pmatrix}, \qquad \mathfrak{B}_2 = \begin{pmatrix} 0 & \epsilon \\ \epsilon^{-1} & 0 \end{pmatrix}$$

leave $2x_1x_2$ invariant. Thus if $f > 1$ the number of units is infinite. If $f = 1$ it is trivial to see that $2x_1x_2$ has only a finite number of units. Theorem 1 is thus completely proved.

Exactly similar considerations hold good if we consider hermitian forms over algebraic number fields. We have only to modify Lemma 3.

LEMMA 4. *Let a and b be integers of K. Then $ax\bar{x} - by\bar{y}$ has an infinity of units in K_0 unless $-ab$ is totally positive in K.*

The method is similar to Lemma 3 and can be easily completed. The analogue of Theorem 1 for hermitian forms is

THEOREM 2. *A hermitian form $\mathfrak{x}'\mathfrak{S}\bar{\mathfrak{x}}$ with integral elements in K_0 has an infinity of units unless $\mathfrak{x}'\mathfrak{S}\bar{\mathfrak{x}}$ is totally definite.*

5. Fundamental domains. Let $\mathfrak{S}$ be the matrix of a non-degenerate quadratic form with integral elements in K. Let H be the space of positive systems $(\mathfrak{H})$ attached to $(\mathfrak{S})$. Let $\mathfrak{B}^{(r)}$ be a real non-singular matrix leaving $\mathfrak{S}^{(r)}$ invariant. Consider the system

$$(17) \qquad (\mathfrak{B}) = (\mathfrak{B}^{(1)}, \cdots, \mathfrak{B}^{(f)}).$$

The systems $(\mathfrak{B})$ form a multiplicative group—multiplication being component wise: $(\mathfrak{B}_1\mathfrak{B}_2) = (\mathfrak{B}_1{}^{(1)}\mathfrak{B}_2{}^{(1)}, \cdots, \mathfrak{B}_1{}^{(f)}\mathfrak{B}_2{}^{(f)}) = (\mathfrak{B}_1) \cdot (\mathfrak{B}_2)$. Let Ω be this multiplicative group. In fact Ω is a Lie group. If $\mathfrak{U}$ is a unit of $\mathfrak{S}$ then $(\mathfrak{U}) = (\mathfrak{U}^{(1)}, \cdots, \mathfrak{U}^{(f)})$ is an element of Ω. Thus Γ_1, the group of units of $\mathfrak{S}$, is a discrete subgroup of Ω. It can easily be verified that if $(\mathfrak{H})$ is an element of H and $\mathfrak{B} \varepsilon \Omega$ then the system

$$(\mathfrak{B}'\mathfrak{H}\mathfrak{B}) = (\cdots, \mathfrak{B}^{(r)\prime}\mathfrak{H}^{(r)}\mathfrak{B}^{(r)}, \cdots)$$

is also in H. Thus Ω has a representation

$$(18) \qquad (\mathfrak{H}) \rightarrow (\mathfrak{B}'\mathfrak{H}\mathfrak{B})$$

as a group of homeomorphisms of H. If $\mathfrak{B} \varepsilon \Gamma_1$, then (18) gives a representation of Γ_1 in H. This representation is discontinuous in H since the group of unimodular matrices is discontinuous in P. (18) will be a faithful representation if we identify $\mathfrak{B}$ and $-\mathfrak{B}$, since both lead to the same point.

(18) thus gives a faithful and discontinuous representation of Γ where Γ is the group Γ_1 in which $\mathfrak{U}$ and $-\mathfrak{U}$ are identified. We now construct a fundamental domain for (18) in H.

From Humbert's theorem [7] on the finiteness of class number it is evident that in each class of the system $(\mathfrak{S})$ there is only a finite number of reduced systems. Let $(\mathfrak{S}[\mathfrak{U}_1]),\cdots,(\mathfrak{S}[\mathfrak{U}_g])$ be the reduced systems. Then $(\mathfrak{H}[\mathfrak{U}_1]),\cdots,(\mathfrak{H}[\mathfrak{U}_g])$ are reduced for certain $(\mathfrak{H})$ in H. Let $R_k = R(\mathfrak{U}_k^{-1})$ denote the set of points $(\cdots,\mathfrak{T}^{(r)}[\mathfrak{U}_k^{(r)-1}],\cdots)$ where $(\mathfrak{T}^{(1)},\cdots,\mathfrak{T}^{(f)})$ is a point of R. Let $H \cap R_k = H_k$ and $H_1 + \cdots + H_g = G$. Then we can prove by exactly the same methods as in [8] that G does not contain in its interior (relative to H) two points equivalent under Γ. The same method shows that equivalent points on the boundary can only be finite in number. Thus G is bounded by a finite number of algebraic surfaces. G has only a finite number of neighbours $G_1,\cdots,G_k$ meaning thereby that if $G(\mathfrak{U})$ denote the image of G under the mapping $(\mathfrak{H}) \rightarrow (\mathfrak{U}'\mathfrak{H}\mathfrak{U})$, $\mathfrak{U} \varepsilon \Gamma$, then G and $G(\mathfrak{U})$ intersect in a non-null set only for $G(\mathfrak{U}) = G_1,\cdots,G_k$. Here we consider G as a closed set, closed relative to H. This would mean of course that if $\mathfrak{U}$ runs through all elements in Γ then $G(\mathfrak{U})$ gives a complete covering of H without gaps, there being overlapping only on the boundary. If in the above, $G_i = G(\mathfrak{U}_i)$, $\mathfrak{U}_i \varepsilon \Gamma$, then from the fact that H is a connected point set it follows that $\mathfrak{U}_1,\mathfrak{U}_2,\cdots,\mathfrak{U}_k$ form a set of generators of Γ. Hence the

THEOREM 1′. *The group Γ_1 of units of $\mathfrak{S}$ is finitely generated.*

Theorem A stated in Section 1 is thus completely proved.

In the case of hermitian forms also we may show that the group of units is finitely generated. To get a faithful representation of Γ_2, the group of units, we must identify $\mathfrak{U}$ and $\omega\mathfrak{U}$ where ω is a root of unity in K_0. (The roots of unity in K_0 form a finite group.)

6. Convergence of the volume integral. We shall now obtain a representation for Γ in the space of the parameters $y_{kl}^{(r)}$, $k=1,\cdots,m_r$; $l=1,\cdots,n_r$; $r=1,\cdots,f$. We define the space J to be the open subspace

$$\mathfrak{T}^{(r)} = \mathfrak{S}^{(r)-1}\begin{bmatrix}\mathcal{E}\\ \mathfrak{Y}^{(r)}\end{bmatrix} > 0, \qquad (r=1,\cdots,f), \tag{19}$$

of a space of M real dimensions. The transformation (4) maps the space J *upon* H. That the mapping (4) is upon may be deduced from the fact that (4) is a parametric representation of H or by following the method in [9].

Let

$$\mathfrak{V}'_r = \begin{pmatrix} \mathfrak{D}^{(r)} \mathfrak{C}^{(r)} \\ \mathfrak{B}^{(r)} \mathfrak{A}^{(r)} \end{pmatrix}$$

and $\mathfrak{V}'_r \mathfrak{S}^{(r)} \mathfrak{V}_r = \mathfrak{S}^{(r)}$, so that $(\mathfrak{V}) = (\mathfrak{V}_1, \cdots, \mathfrak{V}_f)$ is an element of Ω. Then it is easy to see that the transformation

$$(20) \qquad (\mathfrak{Y}) \rightarrow (\cdots, (\mathfrak{A}^{(r)} \mathfrak{Y}^{(r)} + \mathfrak{B}^{(r)}) (\mathfrak{C}^{(r)} \mathfrak{Y}^{(r)} + \mathfrak{D}^{(r)})^{-1}, \cdots)$$

leaves the space J invariant. J is a space of M variables and (20) is a linear transformation of J into itself. (20) is thus a representation of Ω in J. If $\mathfrak{V} \varepsilon \Gamma$, then (20) gives a faithful and discontinuous representation of Γ in J. We can therefore, in J, construct a fundamental domain F bounded by a finite number of hyperplanes.

J is a symmetric riemannian space in the sense of E. Cartan. We can introduce easily a volume measure in this space. For let

$$(21) \qquad (\mathfrak{Y}_1) = (\cdots, (\mathfrak{A}^{(r)} \mathfrak{Y}^{(r)} + \mathfrak{B}^{(r)}) (\mathfrak{C}^{(r)} \mathfrak{Y}^{(r)} + \mathfrak{D}^{(r)})^{-1}, \cdots);$$

$$\mathfrak{S}_1{}^{(r)} = \mathfrak{V}_r' \mathfrak{S}^{(r)} \mathfrak{V}_r, \qquad (r = 1, \cdots, f),$$

where $\mathfrak{V}_r$ is non-singular and $(\mathfrak{V}) = (\mathfrak{V}_1, \cdots, \mathfrak{V}_f)$ need not necessarily be in Ω. It is easy to see that the volume element

$$(22) \qquad dv = N \mid \mathfrak{S}^{(k)} \mid^{-n_k/2} N \mid \mathfrak{T}^{(k)} \mid^{-m/2} d\mathfrak{Y}^{(1)} \cdots d\mathfrak{Y}^{(f)},$$

where

$$(23) \qquad d\mathfrak{Y}^{(r)} = \prod_{k=1}^{m_r} \prod_{l=1}^{n_r} dy_{kl}{}^{(r)},$$

so that $d\mathfrak{Y}^{(1)} \cdots d\mathfrak{Y}^{(f)}$ is the Euclidean volume element in J, remains invariant under (21). The volume element (22) has another interesting property. Of the pairs of integers n_r, m_r, $(r = 1, \cdots, f)$ there are some $n_r \leq m_r$. Let us assume, without loss of generality, that

$$(24) \qquad n_r \leq m_r, \quad (r = 1, \cdots, s); \qquad n_r > m_r, \quad (r = s + 1, \cdots, f).$$

Let α be an integer in K such that

$$(25) \qquad \alpha^{(i)} > 0, \quad (i = 1, \cdots, s); \qquad \alpha^{(i)} < 0, \quad (i = s + 1, \cdots, f).$$

That such integers α exist in K can easily be seen; for if $\omega_1, \cdots, \omega_f$ is a basis of numbers of K consider the inequalities

$$(26) \qquad \sum_{k=1}^{f} \beta_k \omega_k{}^{(i)} > 0, \quad (i = 1, \cdots, s); \qquad < 0, \quad (i = s + 1, \cdots, f).$$

These f inequalities define f open half spaces in f dimensional Euclidean space. The intersection of these f half spaces is an open unbounded set and hence (26) has rational solutions $\beta_1, \cdots, \beta_f$.

Consider now the system $(\alpha\mathfrak{S})$ where α is an integer defined in (25). Then $(\mathfrak{S})$ and $(\alpha\mathfrak{S})$ have the same group of units. We shall show that the volume element (23) remains invariant if $(\mathfrak{S})$ is replaced by $(\alpha\mathfrak{S})$ and at the same time we interchange $n_r, m_r, (r = s+1, \cdots, f)$. Now let $\mathfrak{B}_r = (\alpha^{(r)})^{1/2}\mathcal{E}$, $(r = 1, \cdots, s)$ and $\mathfrak{B}_r = (-\alpha^{(r)})^{1/2}\mathcal{E}$, $(r = s+1, \cdots, f)$. Then by (21) the volume element remains unchanged. We have therefore only to consider invariance under changing $\mathfrak{S}^{(r)}$ into $-\mathfrak{S}^{(r)}$, $(r = s+1, \cdots, f)$, and simultaneously interchanging $n_r, m_r, (r = s+1, \cdots, f)$. This is but an immediate consequence of a corresponding theorem of Siegel's [8].

Let us call a binary quadratic form (hermitian form) *decomposable* if by a linear transformation in K (in K_0) its matrix can be brought to the form

$$\begin{pmatrix} 0 & a \\ a & b \end{pmatrix}; \qquad \begin{pmatrix} 0 & \bar{a} \\ a & b \end{pmatrix}, b = \bar{b}.$$

We shall now prove the

THEOREM B. *Let $\mathfrak{S}$ be not the matrix of a decomposable binary quadratic form; then the volume of a fundamental domain F in J is finite. In other words*

$$N \mid \mathfrak{S}^{(k)} \mid^{-n_k/2} \int_F N \mid \mathfrak{T}^{(k)} \mid^{-m/2} d\mathfrak{Y}^{(1)} \cdots d\mathfrak{Y}^{(f)} \tag{27}$$

converges.

We have made an exception of binary decomposable forms. In Section 7 we shall show that in this case the integral (27) diverges.

The integral in (27) gives a measure of the group of units, Γ. This measure is the same for two isomorphic groups. Since the units of $\mathfrak{S}$ and $\alpha\mathfrak{S}$ are isomorphic, their unit groups have the same measure. We shall therefore, in our proof, assume that the system of signatures of $\mathfrak{S}$ itself satisfies

$$n_r \leq m_r, \qquad (r = 1, \cdots, f). \tag{28}$$

Furthermore, since two matrices in a class have their unit groups isomorphic, it follows that (27) depends only on the class of the system $(\mathfrak{S})$. Theorem B would have followed very simply from the properties of measures provided F was compact (relative to J). But this is not in general the case.

We formulate the corresponding theorem in the case of hermitian forms. It is easy to see that the invariant volume element is[5]

---

[5] The volume element (29) can be computed thus. Let $\mathfrak{Z}_1 = (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$. Let $\delta\mathfrak{Z}$ denote the matrix of differentials $\delta x_{kl} + i\delta y_{kl}$. Then $(\delta z'_{kl}) = \delta\mathfrak{Z}_1 = (\mathfrak{A} - \mathfrak{Z}_1\mathfrak{C})$

$$(29)\qquad dv = 2^{M_1 N} \mid \mathfrak{S}^{(k)} \mid^{-nk/2 N} \mid \mathfrak{T}^{(k)} \mid^{-m/2} d\mathfrak{Z}^{(1)} \cdots d\mathfrak{Z}^{(f)},$$

where $\mathfrak{Z}^{(r)}$ is a complex matrix of m_r rows and n_r columns. Thus

$$(30)\qquad \mathfrak{Z}^{(r)} = \mathfrak{X}^{(r)} + i\mathfrak{Y}^{(r)}; \quad \mathfrak{X}^{(r)} = (x_{kl}{}^{(r)}), \quad \mathfrak{Y}^{(r)} = (y_{kl}{}^{(r)}), \quad i^2 = -1;$$
$$d\mathfrak{Z}^{(r)} = \prod_{k,l} dx_{kl}{}^{(r)} dy_{kl}{}^{(r)}.$$

THEOREM B′. *If F is a fundamental domain for the transformations*

$$(\mathfrak{Z}) \rightarrow (\cdots, (\mathfrak{A}^{(r)}\mathfrak{Z}^{(r)} + \mathfrak{B}^{(r)})(\mathfrak{C}^{(r)}\mathfrak{Z}^{(r)} + \mathfrak{D}^{(r)})^{-1}, \cdots)$$

where $\begin{pmatrix}\mathfrak{D}^{(r)}, \mathfrak{B}^{(r)}\\ \mathfrak{C}^{(r)}, \mathfrak{A}^{(r)}\end{pmatrix}'$ *is a unit of the hermitian form $\mathfrak{x}'\mathfrak{S}\bar{\mathfrak{x}}$ with integral elements in K_0 in the subspace*

$$(31)\qquad (\mathcal{E}\mathfrak{Z}^{(r)\prime})\mathfrak{S}^{(r)-1}\begin{pmatrix}\mathcal{E}\\ \mathfrak{Z}^{(r)}\end{pmatrix} > 0, \qquad (r = 1, \cdots, f),$$

of the space of $M_1 = 2M$ real variables, then

$$(32)\qquad \int_F dv < \infty,$$

where dv is the volume element (29).

It must be noticed that unlike binary decomposable quadratic forms, the units of binary decomposable hermitian forms do *not* have infinite measure.

We shall now prove Theorems B and B′. We shall give in detail the proof of Theorem B. That of B′ is quite similar and we make a few remarks about it at the end of this section.

As pointed out before we shall assume that the system of signatures of $(\mathfrak{S})$ satisfies (28). The transformations (4) show that to every point in J corresponds only one point in H. Let $\mathfrak{U}_1, \cdots, \mathfrak{U}_g$ be the unimodular matrices such that $(\mathfrak{S}[\mathfrak{U}_1]), \cdots, (\mathfrak{S}[\mathfrak{U}_g])$ are reduced. Then for some $\mathfrak{H}$ $(\mathfrak{H}[\mathfrak{U}_1]), \cdots, (\mathfrak{H}[\mathfrak{U}_g])$ are reduced in the sense of Humbert. F is the image in J of a fundamental domain G in H. $G = H_1 + H_2 + \cdots + H_g$, where H_a denotes the set of $(\mathfrak{H})$ in G for which $(\mathfrak{H}[\mathfrak{U}_a])$ is reduced. Also $(\mathfrak{H}[\mathfrak{U}_a])$ satisfies

$$(\mathfrak{H}_a) = (\mathfrak{H}[\mathfrak{U}_a]), \qquad (\mathfrak{S}_a) = (\mathfrak{S}[\mathfrak{U}_a]), \qquad \mathfrak{H}_a{}^{(r)}\mathfrak{S}_a{}^{(r)-1}\mathfrak{H}_a{}^{(r)} = \mathfrak{S}_a{}^{(r)}.$$

$\times\, \delta\mathfrak{Z}(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$. Let $\mathfrak{U} = (u_{kl}) = (\mathfrak{A} - \mathfrak{Z}_1\mathfrak{C})$, $\mathfrak{B} = (v_{kl}) = (\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$. Then $\delta z'_{kl} = \sum_{t,p} u_{kt}\delta z_{tp}v_{pl}$. The Jacobian of the transformation is thus a matrix of order $2mn$. Let $\left|\dfrac{d\mathfrak{Z}_1}{d\mathfrak{Z}}\right|$ denote the absolute value of this Jacobian. Then

$$\left|\frac{d\mathfrak{Z}_1}{d\mathfrak{Z}}\right| = \begin{vmatrix}\mathfrak{W} & \mathfrak{T}\\ -\mathfrak{T} & \mathfrak{W}\end{vmatrix} = \mathrm{abs}\,(\mathfrak{W} + i\mathfrak{T})^2,$$

where $\mathfrak{W} = (Ru_{kt}Rv_{pl} - Iu_{kt}Iv_{pl})$, $\mathfrak{T} = (Iu_{kt}Rv_{pl} + Ru_{kt}Iv_{pl})$ and R and I denote real and imaginary parts.

Let F_a denote that part of F which corresponds to H_a. Then

$$\int_F dv = \sum_{a=1}^{g} \int_{F_a} dv. \tag{33}$$

With every positive system $(\mathfrak{H})$ is associated a unique integral matrix belonging to the set $\mathfrak{A}_1, \cdots, \mathfrak{A}_\mu$. Let $\mathfrak{A}_t$ correspond to the reduced system $(\mathfrak{H}_a)$. Then $(\mathfrak{H}_a[\mathfrak{A}_t])$ is a point in the Humbert space R_0. We subdivide F_a into $F_{a1}, \cdots, F_{at}$, where F_{at} is that part of F_a for which the corresponding $(\mathfrak{H}_a)$ which are reduced are attached to $\mathfrak{A}_t$. Using the invariance of the volume element under (21) we get

$$\int_{F_a} dv = \sum_{i=1}^{t} \int_{F_{at}} dv. \tag{34}$$

The $(\mathfrak{H}_a[\mathfrak{A}_t])$ are points of the Humbert space R_0. Also $(\mathfrak{S}_a[\mathfrak{A}_t])$ are integral matrices and if we write $(\mathfrak{H}) = (\mathfrak{H}_a[\mathfrak{A}_t])$, $(\mathfrak{S}) = (\mathfrak{S}_a[\mathfrak{A}_t])$, then $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$. To prove Theorem B it is necessary to prove $\int_{F_{at}} dv < \infty$ for all a and t. Consider the points $(\mathfrak{H})$ satisfying

$$0 < h_k^{(\lambda)} \leq c_1 h_l^{(\mu)}, \qquad k \leq l, \tag{35}$$

$$\operatorname{abs} h_{kl}^{(\lambda)} \leq c_2 h_k^{(\lambda)}, \qquad k \leq l, \tag{36}$$

$$h_1^{(\lambda)} \cdots h_m^{(\lambda)} \leq c_3 \mid \mathfrak{H}^{(\lambda)} \mid, \tag{37}$$

where $\mathfrak{H}^{(\lambda)} = (h_{kl}^{(\lambda)})$, and $h_k^{(\lambda)} = h_{kk}^{(\lambda)}$, $(k = 1, \cdots, m)$; c_1, c_2, c_3 are constants depending only on K and m. This point set $R_0(c_1, c_2, c_3)$ is a closed point set called the *Humbert space* $R_0(c_1, c_2, c_3)$, and contains all the points $(\mathfrak{H}[\mathfrak{A}])$ where $(\mathfrak{H})$ is a point in the reduced domain R and $\mathfrak{A}$ is a suitable one of the set $\mathfrak{A}_1, \cdots, \mathfrak{A}_\mu$. Let c_1, c_2, c_3 be chosen large enough, in particular $c_1 > 1$. Consider the point set X_{at} in J such that the points $(\mathfrak{H}_a[\mathfrak{A}_t])$ corresponding to the points in X_{at} lie in $R_0(c_1, c_2, c_3)$. Obviously F_{at} is contained in X_{at}. For the proof of Theorem B it suffices to show that $\int_{X_{at}} dv < \infty$. Because of the invariance of the volume element under (21) it suffices to prove that $\int dv$ for those points in the J space for which the corresponding $(\mathfrak{H})$ satisfy (35), (36), (37) is finite. The success of the proof depends largely on the fact that $c_1 > 1$ and on the inequality (35) which does not depend on λ or μ.

Since $\mathfrak{H}^{(r)}$ is positive definite, $h_1^{(r)}, \cdots, h_m^{(r)}$ are all positive and it can be easily proved that (if $\mathfrak{S}^{(r)} = (s_{kl}^{(r)})$)

$$\operatorname{abs} s_{kl}^{(r)} \leq (h_k^{(r)} h_l^{(r)})^{1/2}, \qquad (r = 1, \cdots, f). \tag{38}$$

Since $|\mathfrak{S}^{(r)}| \neq 0$, at least one term in its expansion is not zero. Also $\mathfrak{S}^{(r)}$ is an integral matrix so that it is possible to find for each k an $l = l_k$ such that

(39) $$1 \leq N(s_{kl_k}{}^{(r)}) \leq N(h_k{}^{(r)} h_{l_k}{}^{(r)})^{1/2}.$$

Since $N(h_k{}^{(r)} h_{l_k}{}^{(r)}) \geq 1$, there is a μ depending on k and r such that

(40) $$h_k{}^{(\mu)} h_{l_k}{}^{(\mu)} \geq c_1{}^{-2},$$

for otherwise $N(h_k{}^{(r)} h_{l_k}{}^{(r)}) < c_1{}^{-2f} < 1$, which is false.

Let us fix this μ and let a run over the integers $1, 2, \cdots, m$. Then of the a numbers $l_1, l_2, \cdots, l_a$ at least one of them, say l_k is $\leq m - a + 1$. Hence

(41) $$k \leq a; \qquad l_k \leq m - a + 1.$$

Using (35) with the μ in (40) we get

(42) $$h_a{}^{(r)} \geq c_1{}^{-1} h_k{}^{(\mu)}; \qquad h_{m-a+1}{}^{(r)} \geq c_1{}^{-1} h_{l_k}{}^{(\mu)}.$$

From (40) and (42) follows

(43) $$h_a{}^{(r)} h_{m-a+1}{}^{(r)} \geqq c_1{}^{-4}, \qquad (r = 1, \cdots, f).$$

This inequality is true for all $a = 1, \cdots, m$.

Consider now the products $h_k{}^{(r)} h_{m-k}{}^{(r)}$; $(k = 1, \cdots, m-1)$. There is an integer q_r $(0 \leqq q_r \leqq [m/2])$ such that

(44) $$h_k{}^{(r)} h_{m-k}{}^{(r)} \geq c_1{}^{-4}, \qquad (q_r < k < m - q_r);$$

(45) $$h_{q_r}{}^{(r)} h_{m-q_r}{}^{(r)} < c_1{}^{-4}.$$

q_r might be zero or $[m/2]$. In the first case we do not consider (45) and in the second case we ignore (44). It should be noted that the integers $q_1, \cdots, q_f$ depend only on the point in the J space.

Let $k \leq q_r$ and $l \leq m - q_r$. Then using (35) and (45) we immediately have

(46) $$h_k{}^{(\lambda)} h_l{}^{(\lambda)} < c_1{}^{-2} \qquad (\text{for all } \lambda),$$

and using (38) we obtain

(47) $$0 \leq N(\operatorname{abs} s_{kl}{}^{(r)}) < c_1{}^{-2f} < 1,$$

which means $s_{kl}{}^{(r)} = 0$. Thus

(48) $$s_{kl}{}^{(r)} = 0 \qquad \text{for all } r,\ k \leq q_r,\ l \leq m - q_r.$$

(48) shows that if $q = \operatorname{Max}(q_1, \cdots, q_f)$, then for all λ, $\mathfrak{S}^{(\lambda)}$ has a decomposition

$$\mathfrak{S}^{(\lambda)} = \begin{bmatrix} \mathfrak{O}, & \mathfrak{O}, & \tilde{\mathfrak{P}}_\lambda' \\ \mathfrak{O}, & \tilde{\mathfrak{N}}_\lambda, & \tilde{\mathfrak{Q}}_\lambda' \\ \tilde{\mathfrak{P}}_\lambda, & \tilde{\mathfrak{Q}}_\lambda, & \tilde{\mathfrak{G}}_\lambda \end{bmatrix}, \tag{49}$$

where $\tilde{\mathfrak{N}}_\lambda$ is $(m-2q)$-rowed symmetric and $\tilde{\mathfrak{G}}_\lambda$ is q-rowed symmetric. We may, without contradiction, restrict ourselves to $\mathfrak{S}^{(r)}$ and write

$$\mathfrak{S}^{(r)} = \begin{pmatrix} \mathfrak{O} & \mathfrak{O} & \mathfrak{P}_r' \\ \mathfrak{O} & \mathfrak{N}_r & \mathfrak{Q}_r' \\ \mathfrak{P}_r & \mathfrak{Q}_r & \mathfrak{G}_r \end{pmatrix}, \tag{50}$$

where $\mathfrak{N}_r$ is $m-2q_r$-rowed symmetric and $\mathfrak{G}_r$ is q_r-rowed symmetric respectively.

The essential point of the inequalities (44) and (45) is to obtain:

$$\begin{array}{lll} & h_1^{(r)}, \cdots, h_{q_r}^{(r)} & \text{are bounded;} \\ (51) & h_{q_r+1}^{(r)}, \cdots, h_{m-q_r}^{(r)} & \text{are bounded together with their reciprocals;} \\ & h_{m-q_r+1}^{(r)}, \cdots, h_m^{(r)} & \text{have their reciprocals bounded,} \end{array}$$

these bounds depending on $|\mathfrak{H}|$, K and m only. To prove them we observe the identity

$$\prod_{k=1}^{m}(h_k^{(r)}h_{m-k+1}^{(r)}) = \prod_{k=1}^{q_r}(\cdot\cdot\cdot)^2 \prod_{k=q_r+1}^{m}(\cdot\cdot\cdot) / \prod_{k=1}^{q_r}(\cdot\cdot\cdot). \tag{52}$$

Now $q_r \leq m - q_r$ so that

$$\prod_{k=1}^{m}(\cdot\cdot\cdot) \geq h_{m-q_r}^{(r)2} \prod_{k=1}^{q_r}(\cdot\cdot\cdot)^2. \tag{53}$$

From this we obtain, using (37) and (43), that

$$h_{m-q_r}^{(r)} \leq c_4, \tag{54}$$

c_4 being a constant depending on $|\mathfrak{H}|$, K and m. Now using (44) we have

$$(h_{m-k}^{(r)})^{-1} \leq c_5 h_k^{(r)}, \qquad (q_r < k < m - q_r). \tag{55}$$

Therefore from (54)

$$(h_{m-k}^{(r)})^{-1} \leq c_6, \qquad (q_r < k < m - q_r). \tag{56}$$

Inequalities (54) and (56) together prove (51).

From now on the proof is as in Siegel's paper [8]. We write down only the essential results which may be proved easily.

$$|\mathfrak{T}^{(r)}|^{-m/2} \leq c_7 \prod_{k=1}^{q_r}(h_k^{(r)})^{m/2}, \qquad q_r \leq n_r. \tag{57}$$

If

$$\mathfrak{Y}^{(r)} = \begin{pmatrix} \mathfrak{Y}_1^{(r)} & \mathfrak{Y}_2^{(r)} \\ \mathfrak{Y}_3^{(r)} & \mathfrak{Y}_4^{(r)} \end{pmatrix},$$

then the elements of the matrices $\mathfrak{Y}_1^{(r)}$, $\mathfrak{Y}_2^{(r)}$, $\mathfrak{Y}_4^{(r)}$ are all bounded in absolute value by quantities depending only on K, m and $|\mathfrak{S}|$. Let

$$\mathfrak{Y}_3^{(r)} = \mathfrak{P}_r(\mathfrak{X}_1^{(r)} + \mathfrak{X}_2^{(r)}) \tag{58}$$

where $\mathfrak{X}_1^{(r)}$ is q_r-rowed symmetric and $\mathfrak{X}_2^{(r)}$ is skew symmetric. Also

$$\operatorname{abs} x_{kl}^{(r)} \leq c_8 (h_l^{(r)})^{-1}, \qquad (1 \leq k \leq l \leq q_r), \tag{59}$$

where $\mathfrak{X}_1^{(r)} = (x_{kl}^{(r)})$ and the elements of $\mathfrak{X}_2^{(r)}$ are bounded.

Each point in X_{at} is associated with a system of numbers $q_1, q_2, \cdots, q_f$, where $0 \leq q_r \leq [m/2]$ and $q_r \leq n_r$. Let $I(q_1, \cdots, q_f)$ be the integral extended over all points of X_{at} which have the same $q_1, \cdots, q_f$. Then

$$\int_X dv = \sum_{q_1=0}^{n_1} \sum_{q_2=0}^{n_2} \cdots \sum_{q_r=0}^{n_r} I(q_1, \cdots, q_f). \tag{60}$$

Now $I(0, 0, \cdots, 0)$ is bounded because then from (57), $|\mathfrak{T}^{(r)}|^{-m/2}$ is bounded for all r and from (58) all the elements of $\mathfrak{Y}^{(r)}$ are bounded so that $I(0, \cdots, 0)$ is an integral over a bounded domain with a bounded integrand. Let us now consider $I(q_1, \cdots, q_f)$, where not all $q_1, \cdots, q_f$ are zero. Let us split up the domain where $q_1, \cdots, q_f$ have a specified system of values not all zero into parallelpipeds

$$\begin{gathered} \exp(-tg_k^{(r)}) < h_k^{(r)}/h_{k+1}^{(r)} \leq \exp(t(1 - g_k^{(r)})), \\ (k = 1, \cdots, q_r - 1); \\ \exp(-tg_{q_r}^{(r)}) < h_{q_r}^{(r)} \leq \exp(t(1 - g_{q_r}^{(r)})), \end{gathered} \tag{61}$$

where $c_1 = e^t > 1$ and $g_k^{(r)}$ for all $r = 1, \cdots, f$ and $k = 1, \cdots, q_r$, runs through all integers $0, 1, 2, \cdots$. What we have done is to split up the domain into a number of disjoint subsets. Let $f_k^{(r)} = g_k^{(r)} + \cdots + g_{q_r}^{(r)}$, $f_0^{(r)} = 0$, $k = 1, \cdots, q_r$. Then

$$c_6 \exp(-tf_k^{(r)}) \leq h_k^{(r)} \leq c_7 \exp(-tf_k^{(r)}). \tag{62}$$

Using (57) and (62) we get

$$|\mathfrak{T}^{(r)}|^{-m/2} \leq c_8 \exp[-\tfrac{1}{2} mt \sum_{k=0}^{q_r} f_k^{(r)}]. \tag{63}$$

The volume of every subdomain has the upper bound

$$c_9 \exp[\sum_{r=1}^{f} \sum_{k=0}^{q_r} tkf_k^{(r)}]. \tag{64}$$

Thus $I(q_1, \cdots, q_f)$ is majorised by

$$c_{10} \sum_{A} \exp(-At), \tag{65}$$

where $A = \sum_{r=1}^{f} \sum_{k=0}^{q_r} (\frac{1}{2}m - k) f_k^{(r)}$ and the summation in $g_1^{(r)}, \cdots$ runs through all non-negative rational integers. (65) is actually a convergent series if $m/2 - k > 0$ and diverges if $m = 2$ and $q_1 = q_2 = \cdots = q_f = 1$. This is the case of a binary decomposable form; for in this case $\mathfrak{S}$ is of the form (50), namely

$$\mathfrak{S}^{(r)} = \begin{pmatrix} 0 & p_r \\ p_r & g_r \end{pmatrix}. \tag{66}$$

In the statement of Theorem B we have excepted this case. Thus Theorem B is completely proved.

In the case of hermitian forms over K_0 the arguments are exactly similar. The difference arises in (57) where we have

$$N \,|\, \mathfrak{T}^{(r)} \,|^{-m/2} \leq c_7 \prod_{r=1}^{f} \prod_{k=1}^{q_r} (h_k^{(r)})^m \tag{67}$$

and so the analogue of the function A in (65) is

$$A = \sum_{r=1}^{f} \sum_{k=0}^{q_r} (m + 1 - 2k) f_k^{(r)}, \tag{68}$$

so that the fundamental domain has finite volume since $m + 1 > 2q_r$. We present an interesting proof of the finiteness in the case of decomposable hermitian forms; that is binary hermitian forms whose matrix is:

$$\mathfrak{S} = \begin{pmatrix} 0 & a \\ \bar{a} & b \end{pmatrix}, \qquad b = \bar{b}, \tag{69}$$

where a and b are integers of K_0. We show that in this case the measure is finite.

Let $\mathfrak{S}$ be a non-singular hermitian integral matrix in (69). If we make the transformation in K_0:

$$x_1 = -2(a - \bar{a})^{-1}(ax + \tfrac{1}{2}by), \qquad y_1 = y, \tag{70}$$

then $\mathfrak{S}$ goes into $\frac{1}{2}(\bar{a} - a)(x\bar{y} - y\bar{x})$ whose matrix is $\frac{1}{2}(\bar{a} - a)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. By (21) we can assume $(\bar{a} - a)/(2i)$ is totally positive, for otherwise we can interchange x and y or multiply by a suitable integer of K. It is easily seen that for proving the theorem it is enough to consider the units of $\frac{1}{2}(\bar{a} - a)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. This matrix we denote by $\mathfrak{S}_0$.

Let $\mathfrak{M} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ be a unit of $\mathfrak{S}_0$ in K_0. Then $\alpha\delta - \beta\gamma = \rho$, where ρ is a root of unity in K_0. The roots of unity in K_0 form a finite group [5] and so by considering those units for which $\alpha\delta - \beta\gamma = 1$ we shall only be considering a subgroup of finite index in the group of units. Let us call this subgroup the *principal group of units.* Let $\mathfrak{M} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$, where $\alpha, \beta, \gamma, \delta$ are integers such that $\alpha\delta - \beta\gamma = 1$, be in the principal group. Then

$$(71) \qquad \alpha\bar{\gamma} = \gamma\bar{\alpha}, \qquad \beta\bar{\delta} = \delta\bar{\beta}, \qquad \alpha\bar{\delta} - \beta\bar{\gamma} = 1.$$

The most general solution of these equations is

$$(72) \qquad \alpha/a = \beta/b = \gamma/c = \delta/d = \rho,$$

where ρ is a complex number of absolute value 1 and a, b, c, d are real. Since $\alpha, \beta, \gamma, \delta$ are integers of K_0 we see that a, b, c, d are integers of the totally real field K. Thus this principal group of units is none other than the restricted Hilbert modular group [2] in K.

The representation of this group is in a space of f complex variables $z_1, \cdots, z_f$ where

$$(73) \qquad \mathfrak{S}^{(r)-1}\left\{ \frac{1}{z_r} \right\} = 2/(\bar{a} - a)(\bar{z}_r - z_r) > 0,$$

which is the space of the upper half planes of the f complex variables z_r. The group of units is represented in this space by

$$(74) \qquad z_r \to \alpha^{(r)} z_r + \beta^{(r)}/\gamma^{(r)} z_r + \delta^{(r)}, \qquad (r = 1, \cdots, f),$$

and the volume element is

$$(75) \qquad dv = \prod_{r=1}^{f} | \mathfrak{T}^{(r)} |^{-2} \, dz_1 \cdots dz_f,$$

where $\mathfrak{T}^{(r)} = 2(\bar{a} - a)^{-1}(\bar{z}_r - z_r)$ and $dz_r = 2dx_r dy_r$, $z_r = x_r + iy_r$. The measure of volume is $dx_1 \cdots dx_f dy_1 \cdots dy_f / y_1^2 \cdots y_f^2$. From the work of Blumenthal [2] it is known that (74) has a fundamental domain bounded by a finite number of surfaces and by Siegel's work [11], this fundamental domain has a finite volume.

7. Decomposable binary forms. In this section we shall show that in the case of a binary decomposable quadratic form a fundamental domain has actually infinite volume. From the considerations of Section **4** it is enough if we consider the form $\mathfrak{x}'\mathfrak{S}\mathfrak{x}$ where

$$(76) \qquad \mathfrak{S} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The system $(\mathfrak{S})$ is $(\mathfrak{S}^{(1)}, \cdots, \mathfrak{S}^{(f)})$, where $\mathfrak{S}^{(r)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Let $\mathfrak{U} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ be a unit of $\mathfrak{S}$. It is then easy to see that $\mathfrak{U}$ has one of the two forms

$$(77) \qquad \begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon^{-1} \end{pmatrix}, \quad \begin{pmatrix} 0 & \epsilon \\ \epsilon^{-1} & 0 \end{pmatrix},$$

where ϵ is an algebraic unit in K. Let Γ_0 be the group of units of $\mathfrak{S}$ and Γ_1 the subgroup of units of the form $\begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon^{-1} \end{pmatrix}$. Then Γ_0/Γ_1 is of order 2.

Now the space J can be easily seen to be the space of f real variables $u_1, \cdots, u_f$ for which

$$(78) \qquad u_i > 0, \qquad (i = 1, \cdots, f).$$

The representation of Γ_1 in this space is

$$(79) \qquad u_i \rightarrow \epsilon^{(i)2} u_i, \qquad (i = 1, \cdots, f).$$

The volumes of the fundamental domains for Γ_0 and Γ_1 are in the ratio 1:2, so that it is enough if we consider Γ_1.

From Dirichlet's theorem we see that any unit ϵ may be written as $\epsilon = \pm \epsilon_1^{a_1} \cdots \epsilon_{f-1}^{a_{f-1}}$. Let us make the transformation

$$(80) \qquad \log u_i = v_i, \qquad (i = 1, \cdots, f)$$

(since $u_i > 0$ this is permissible). We have then the whole of the space V of $v_1, \cdots, v_f$. The transformations (79) become

$$(81) \qquad v_i \rightarrow v_i + a_1 \lambda_1^{(i)} + \cdots + a_{f-1} \lambda_{f-1}^{(i)},$$

where $\lambda_j^{(i)} = \log(\operatorname{abs} \epsilon_j^{(i)})^2$, $i = 1, \cdots, f$; $j = 1, \cdots, f-1$. Also since $N\epsilon^2 = 1$ we have for any unit ϵ_j

$$(82) \qquad \sum_{i=1}^{f} \lambda_j^{(i)} = 0,$$

which means that the mappings (81) preserve the planes

$$(83) \qquad v_1 + \cdots + v_f = \text{a constant}.$$

It is known [5] that in an algebraic number field K the matrix

$$(84) \qquad (\lambda_j^{(i)}), \qquad i = 1, \cdots, f;\ j = 1, \cdots, f-1,$$

is of rank $f - 1$. (This is none other than the assertion that the regulator is non-vanishing.) This means that every point $v_1, \cdots, v_f$ in V may be written

$$(85) \qquad v_i = \eta_0/f + \eta_1\lambda_1^{(i)} + \cdots + \eta_{f-1}\lambda_{f-1}^{(i)},$$

where $\eta_0 = v_1 + \cdots + v_f$ and $\eta_1, \cdots, \eta_{f-1}$ are real numbers. Consider now the plane $v_1 + \cdots + v_f = 0$. Because of (81) the points $(\eta_i\lambda_i^{(1)}, \cdots, \eta_i\lambda_i^{(f)})$, $i = 1, \cdots, f-1$, lie in this plane. Consider rectangular coordinates in this plane with origin as centre. Consider the $f-1$ points

$$(86) \qquad (\lambda_i^{(1)}, \cdots, \lambda_i^{(f)}), \qquad i = 1, \cdots, f-1.$$

Because of (82) these points together with the origin form the vertices of a $(f-1)$ simplex. This simplex can be completed into an $(f-1)$ dimensional parallelepiped. We make this a fundamental parallelepiped by prescribing that the boundary points lying on the faces of this parallelepiped containing the origin shall belong to the fundamental region. We call this F_1. Choose now a fixed direction through the origin and not lying in this plane. Through every point of F_1 draw lines parallel to this direction. The resulting point set F can easily be verified by the method of Blumenthal [2] to be a fundamental domain for (79).

The volume element in the $u_1, \cdots, u_f$ space is $du_1 \cdots du_f/u_1 \cdots u_f$. That in the $v_1, \cdots, v_f$ space is the Euclidean volume element $dv_1 \cdots dv_f$. The Euclidean volume of F is obviously infinite. Thus the

THEOREM C. *The volume of a fundamental domain for the units of a binary decomposable quadratic form over a totally real field is infinite.*

In the case of rational decomposable binary forms this theorem is at once obvious from Theorem D below, because the number of units of a binary decomposable form over the field of rationals is finite. In the case of algebraic number fields this result does not follow from Theorem D.

8. Volume of the space *J*.

In this section we shall prove

THEOREM D. *The volume of the space J measured with the volume element* (22) *is infinite.*

This is also true of the space (31) in case of hermitian forms. We shall prove it in the case $f = 1$ only. The general result follows at once by taking topological products.

It is evident that we can assume that $\mathfrak{S} = \begin{pmatrix} \mathfrak{E}_n & \mathfrak{D} \\ \mathfrak{D} & -\mathfrak{E}_{m-n} \end{pmatrix}$. Let us put $n = p$ and $m - n = q$ so that $p + q = m$. The J space is now

$$\mathcal{E} - \mathfrak{Y}'\mathfrak{Y} > 0, \tag{87}$$

where $\mathcal{E}$ is the unit matrix of order p and $\mathfrak{Y}$ is a real matrix of p columns and q rows. The invariant volume element is

$$dv = | \mathcal{E} - \mathfrak{Y}'\mathfrak{Y} |^{-\frac{1}{2}(p+q)} d\mathfrak{Y}. \tag{88}$$

Before proving Theorem D we make the following remarks.

If $\mathfrak{S}$ is a positive definite symmetric matrix and $\mathfrak{S} = \begin{pmatrix} \mathfrak{S}_1 & \mathfrak{S}_2 \\ \mathfrak{S}'_2 & \mathfrak{S}_3 \end{pmatrix}$, where $\mathfrak{S}_1$, $\mathfrak{S}_3$ are symmetric, then

$$| \mathfrak{S} | \leq | \mathfrak{S}_1 | \, | \mathfrak{S}_3 | . \tag{89}$$

In the second place if $\mathcal{E} - \mathfrak{Y}'\mathfrak{Y} > 0$ then $| \mathcal{E} - \mathfrak{Y}'\mathfrak{Y} | \leq 1$. This is easy, for if $\mathfrak{Y} = (y_{kl})$, then the diagonal elements of $\mathcal{E} - \mathfrak{Y}'\mathfrak{Y}$ are

$$1 - \sum_{t=1}^{q} y_{tk}^2 > 0 \qquad (k = 1, \cdots, p).$$

Also these diagonal elements are < 1. From (89) follows the result.

We prove Theorem D by induction on p. When $p = 1$, the space is

$$J_1\colon\ 1 - y_1^2 - \cdots - y_q^2 > 0 \tag{90}$$

which is the interior of the unit sphere in the q dimensional space. It is then well known that

$$\int_{J_1} (1 - y_1^2 - \cdots - y_q^2)^{-\frac{1}{2}(q+1)} dy_1 \cdots dy_q \tag{91}$$

diverges for $q > 0$. Theorem D is thus true for $p = 1$. Let us assume it proved for $p - 1$ instead of p. Let $\mathfrak{Y}_1$ be the first $p - 1$ columns of $\mathfrak{Y}$ and $\mathfrak{x}$ the last column so that $\mathfrak{Y} = (\mathfrak{Y}_1 \mathfrak{x})$. Then

$$\mathcal{E} - \mathfrak{Y}'\mathfrak{Y} = \begin{pmatrix} \mathcal{E} - \mathfrak{Y}_1'\mathfrak{Y}_1 & -\mathfrak{Y}_1'\mathfrak{x} \\ -\mathfrak{x}'\mathfrak{Y}_1 & 1 - \mathfrak{x}'\mathfrak{x} \end{pmatrix} > 0, \tag{92}$$

which shows that $\mathcal{E} - \mathfrak{Y}_1'\mathfrak{Y}_1 > 0$, $1 - \mathfrak{x}'\mathfrak{x} > 0$. Using (89) we get

$$| \mathcal{E} - \mathfrak{Y}'\mathfrak{Y} | \leq | \mathcal{E} - \mathfrak{Y}_1'\mathfrak{Y}_1 | \, (1 - \mathfrak{x}'\mathfrak{x}) \tag{93}$$

and hence

$$\begin{aligned} &\int_{J_p} | \mathcal{E} - \mathfrak{Y}'\mathfrak{Y} |^{-\frac{1}{2}(p+q)} d\mathfrak{Y} \\ &\qquad \geq \int_{J_p} | \mathcal{E} - \mathfrak{Y}_1'\mathfrak{Y}_1 |^{-\frac{1}{2}(p+q-1)} (1 - \mathfrak{x}'\mathfrak{x})^{-(p+1)/2} d\mathfrak{Y}_1 d\mathfrak{x}. \end{aligned} \tag{94}$$

If $\mathfrak{Y}_1$ is a real matrix of $p-1$ columns and q rows, then it is possible to choose a column vector $\mathfrak{x}$ such that

$$(95) \qquad \mathfrak{E}-\mathfrak{Y}'\mathfrak{Y} > 0, \qquad \mathfrak{E}-\mathfrak{Y}_1'\mathfrak{Y}_1 > 0, \qquad \mathfrak{Y} = (\mathfrak{Y}_1\mathfrak{x}).$$

This is seen from the identity

$$(96) \qquad (\mathfrak{E}-\mathfrak{Y}'\mathfrak{Y}) = \begin{pmatrix} \mathfrak{E}-\mathfrak{Y}_1'\mathfrak{Y}_1 & \mathfrak{N} \\ \mathfrak{N} & \mathfrak{Q} \end{pmatrix} \begin{bmatrix} \mathfrak{E} & -(\mathfrak{E}-\mathfrak{Y}_1'\mathfrak{Y}_1)^{-1}\mathfrak{Y}_1'\mathfrak{x} \\ \mathfrak{N} & 1 \end{bmatrix},$$

where

$$(97) \qquad \mathfrak{Q} = 1 - \mathfrak{x}'\mathfrak{x} - \mathfrak{x}'\mathfrak{Y}_1(\mathfrak{E}-\mathfrak{Y}_1'\mathfrak{Y}_1)^{-1}\mathfrak{Y}_1'\mathfrak{x} > 0.$$

Now (97) is necessary and sufficient for $\mathfrak{E}-\mathfrak{Y}'\mathfrak{Y} > 0$. Also (96) implies that $1-\mathfrak{x}'\mathfrak{x} > 0$, so that if J_{p-1} is the space

$$(98) \qquad \mathfrak{E}-\mathfrak{Y}_1'\mathfrak{Y}_1 > 0,$$

then from (94) we get

$$(99) \qquad \int_{J_p} |\,\mathfrak{E}-\mathfrak{Y}'\mathfrak{Y}\,|^{-(p+q)/2} d\mathfrak{Y} \geq \int_{J_{p-1}} |\,\mathfrak{E}-\mathfrak{Y}_1'\mathfrak{Y}_1\,|^{-\frac{1}{2}(p+q-1)} d\mathfrak{Y}_1,$$

which now completes the induction.

It is possible to obtain Theorem D from topological considerations alone. Siegel [10] has shown that the space H of positive solutions of

$$(100) \qquad \mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$$

can be considered as the homogeneous space of the cosets of a Lie group $\mathfrak{G}$ by a closed maximal compact subgroup $\mathfrak{C}$. From the existence of invariant measures in $\mathfrak{G}$ and $\mathfrak{G}/\mathfrak{C}$ it is evident that $\mathfrak{G}/\mathfrak{C}$ has infinite measure for otherwise, $\mathfrak{C}$ being compact, $\mathfrak{G}$ has finite measure and therefore is compact, which is absurd. Theorem D is also implicit in Theorem 1 of Siegel [10]. For we can show that H contains a geodesic sphere of arbitrarily large radius and since the volume of a geodesic sphere of radius r is $> v(r)$ where $v(r)$ is the volume of a Euclidean sphere of pq dimensions and of radius r, the theorem D becomes obvious.

INSTITUTE FOR ADVANCED STUDY,
PRINCETON, NEW JERSEY.

BIBLIOGRAPHY.

[1] E. Artin, *Galois Theory*, Notre Dame, 1944.

[2] O. Blumenthal, "Über Modulfunktionen von mehreren Veränderlichen," *Mathematische Annalen*, Bd. 56 (1903), pp. 509-548.

[3] Dirichlet-Dedekind, *Vorlesungen über Zahlentheorie*, Vierte Auflage, p. 149.

[4] R. Fricke, "Zur gruppentheoretische Grundlegung der automorphen Funktionen," *Mathematische Annalen*, vol. 42 (1893), pp. 564-594.

[5] E. Hecke, *Theorie der algebraischen Zahlen*, Leipzig, 1923.

[6] P. Humbert, "Théorie de la réduction des formes quadratiques défines positives dans un corps algébrique K fini," *Commentarii Mathematici Helvetici*, vol. 12 (1940), pp. 263-306.

[7] ———, "Réduction des formes quadratiques dans un corps algébrique fini," *ibid.*, vol. 23 (1949), pp. 50-63.

[8] C. L. Siegel, "Einheiten quadratischer Formen," *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität*, vol. 13 (1940), pp. 209-239.

[9] ———, "On the theory of indefinite quadratic forms," *Annals of Mathematics*, vol. 45 (1944), pp. 577-622.

[10] ———, "Some remarks on discontinuous groups," *ibid.*, vol. 46 (1945), pp. 708-718.

[11] ———, "The volume of the fundamental domain for some infinite groups," *Transactions of the American Mathematical Society*, vol. 39 (1936), pp. 209-218.

THE EXISTENCE AND LIMIT BEHAVIOR OF THE ONE-DIMENSIONAL SHOCK LAYER.*

By David Gilbarg.

Introduction. We consider steady, one-dimensional flows of a viscous, heat-conducting fluid which approach finite limit values at $x = +\infty$ and $x = -\infty$. Such flows display the character of a shock wave (for small viscosity, μ, and heat conductivity, λ), in that they differ sensibly from their end states at $x = \pm\infty$ only in a small interval of rapid transition. In analogy with the classical boundary layer, and also to distinguish these flows from the shock waves which belong properly to the theory of ideal fluids, we follow Weyl [1] in naming such a flow a *shock layer*. The one-dimensional shock layer is in certain respects the prototype of all shock phenomena and has therefore been studied widely, with particular emphasis on the problem of thickness of the shock front [2, 3, 5, 6, 8]. However, basic problems concerning these flows, such as those of existence, and limit behavior for small λ, μ, remain open. Their solution, which we consider here, is a step towards placing on a sound basis the relation between the theories of real and ideal fluids.

The general problem of existence of the shock layer for a fluid with given λ, μ, and with the preassigned end states, has been studied inconclusively by Rayleigh [4] and Weyl [1]. Until now, the existence of the shock layer seems to have been definitely proved only for an exceptional set of ideal gases for which a postulated relation between λ, μ, and the specific heat at constant pressure,[1] permits explicit integration of the equations of motion; (Becker [2], also [5, 6]). We succeed here in obtaining an essentially complete solution of the existence problem by proving the existence and uniqueness of the shock layer for the general class of fluids considered by Weyl, with λ, μ arbitrary functions of the state, and for arbitrary end states satisfying the shock relations (Theorem 1). This result, therefore, establishes for general fluids an exact correspondence between the steady one-dimensional shock waves and the shock layers.

---

* Received June 3, 1950; revised November 13, 1950.

Prepared under Navy Contract N6onr-180, Task Order V, with Indiana University.

[1] Namely, Prandtl number $c_p\mu/\lambda = 3/4$.

The problem of the limit behavior of the shock layers for small λ and μ is that of providing a rigorous proof for the statement, generally accepted on physical grounds, that any shock wave in an ideal fluid is the limit of the corresponding shock layers, and, conversely, that the shock layers (with the same end states) approach a shock wave in the limit as $\lambda, \mu \rightarrow 0$.[2] Mathematically, this problem belongs to the class of perturbation problems for differential equations in which the vanishing of certain parameters, (λ, μ in this case), reduces the order of the equations. A complete solution to the problem is given in Theorems 2 and 5 for the same general fluids as in the existence theorem. The limit behavior makes apparent the unequal dependence of the flows on viscosity and heat conductivity in the fact that as $\lambda \rightarrow 0$ (with μ fixed), the shock layers converge to a continuous thermally non-conducting shock layer (Theorem 3), whereas for $\mu \rightarrow 0$ (with λ fixed), they converge to a (generally) *discontinuous* non-viscous shock layer (Theorem 4).

We remark that the methods used here provide easily estimated bounds on the thickness of the shock front, although not the precise determinations sought by previous writers [2, 3, 5, 6] on this subject.

1. **Existence and uniqueness of the shock layer.** The equations of motion for a steady one-dimensional flow of a viscous heat conducting fluid are:[3]

$$\rho u = \text{constant} = c_1; \tag{1}$$

$$\rho u^2 + p - \mu u_x = \text{constant} = c_2 \tag{2}$$

$$\rho u(\tfrac{1}{2}u^2 + e + p/\rho) - \mu u u_x - \lambda \theta_x = \text{const.} = c_3. \tag{3}$$

These equations express respectively the conservation conditions for mass, momentum, and energy. The coefficient of viscosity, μ (which combines the two viscosity coefficients appearing in the general Navier-Stokes equations), and the heat conductivity, λ, are in general functions of the thermodynamic state, and will here be considered in this generality. The other quantities appearing in the equations are the density ρ, pressure p, velocity u, internal energy e per unit mass, and temperature θ. We assume that the variables θ and $\tau = 1/\rho =$ specific volume, fix uniquely the thermodynamic state of the fluid, and that λ, μ, p, and e are sufficiently smooth (e. g. twice differentiable), functions of these variables.

---

[2] This limit behavior is apparent in the explicit solution obtained by Becker [2]; however, in this case, λ and μ are connected by the relation, $c_p\mu/\lambda = 3/4$, and do not approach zero independently. For discussion of the problem of limit behavior, see, for example, [7] pp. 135-138, also [8], pp. 218-222.

[3] See, for example, [7] p. 134; a complete derivation is contained in [9].

Eliminating u from equations (1)-(3), we obtain,

$$\text{(E)}\quad \begin{aligned} \lambda(d\theta/dx) &= b[e(\tau,\theta) - \tfrac{1}{2}b^2(\tau - a)^2 - c] \equiv L(\tau,\theta) \\ \mu(d\tau/dx) &= 1/b[p(\tau,\theta) + b^2(\tau - a)] \equiv M(\tau,\theta), \end{aligned}$$

where,

$$a = c_2/c_1^2, \qquad b = c_1, \qquad c = c_3/c_1 - c_2^2/2c_1^2,$$

and $e = e(\tau,\theta)$, $p = p(\tau,\theta)$ are (given) equations of state of the fluid. With certain minor modifications, this is the form of the equations of motion used by both Weyl [1] and Becker [2].

We enumerate now a set of conditions on the functions $L(\tau,\theta)$, $M(\tau,\theta)$ which will be sufficient to prove the desired existence and limit theorems. In Section **3**, it will be verified that the corresponding functions for the general fluids of Weyl, and, in particular, for polytropic gases, satisfy these conditions. We assume:

The domain of definition is a set of points $Z = (\tau,\theta)$, hereafter called the Z-plane, forming a simply connected region in the quadrant $\tau > 0$, $\theta > 0$. In this region, we have:

(A) $L_\theta > 0$, $M_\theta > 0$.

(B) There are two curves, L and M, on which $L(\tau,\theta) = 0$ and $M(\tau,\theta) = 0$, respectively, and which intersect in two points, $Z_0 = (\tau_0,\theta_0)$, $Z_1 = (\tau_1,\theta_1)$, $(\tau_0 > \tau_1)$; (τ_0,θ_0) and (τ_1,θ_1) are the only simultaneous solutions of $L(\tau,\theta) = 0$ and $M(\tau,\theta) = 0$.

(C) $L_\tau > 0$ on L for $\tau_1 \leqq \tau \leqq \tau_0$.

(D) $L_\tau/L_\theta < M_\tau/M_\theta$ at Z_0; $L_\tau/L_\theta > M_\tau/M_\theta$ at Z_1.

It is an immediate consequence of (A) that L and M are representable as single valued functions of τ, namely, $\theta = l(\tau)$, $\theta = m(\tau)$, respectively. By virtue of (C), $l(\tau)$ is monotonically decreasing in the interval $\tau_1 \leqq \tau \leqq \tau_0$, and thus $\theta_1 > \theta_0$. From (B) and (D) it follows $m(\tau) > l(\tau)$ for $\tau_1 < \tau < \tau_0$.

The closed curve formed by the arcs $L^* \subset L$ and $M^* \subset M$ which join Z_0 and Z_1 bounds a simply connected region R of the Z-plane (Fig. 1). We conclude from (A) that everywhere in R, $L(\tau,\theta) > 0$ and $M(\tau,\theta) < 0$.

The preceding geometric facts contain all the information that is needed about the (τ,θ) phase plane for the proofs which follow. More general conditions could be obtained, but would be of no interest for the ensuing applications.

Consider the points $Z_0(\tau_0,\theta_0)$ and $Z_1(\tau_1,\theta_1)$. These satisfy the shock conditions,

$$p_0 + b^2\tau_0 = p_1 + b^2\tau_1 = b^2 a \tag{4}$$

$$e_0 - b^2/2(\tau_0 - a)^2 = e_1 - b^2/2(\tau_1 - a)^2 = c. \tag{5}$$

Z_0 and Z_1 then represent possible initial and final states, respectively, of a normal shock wave of an ideal fluid having the same equations of state as the given fluid. A solution, $S(x) = (\tau(x), \theta(x))$, $(-\infty < x < +\infty)$, of equations (E), (for given $\lambda(\tau, \theta)$, $\mu(\tau, \theta)$), will be called a *shock layer* if

$$\lim_{x \to -\infty} S(x) = Z_0, \quad \lim_{x \to -\infty} S(x) = Z_1.$$

The corresponding *shock layer curve* in the Z-plane is the integral curve

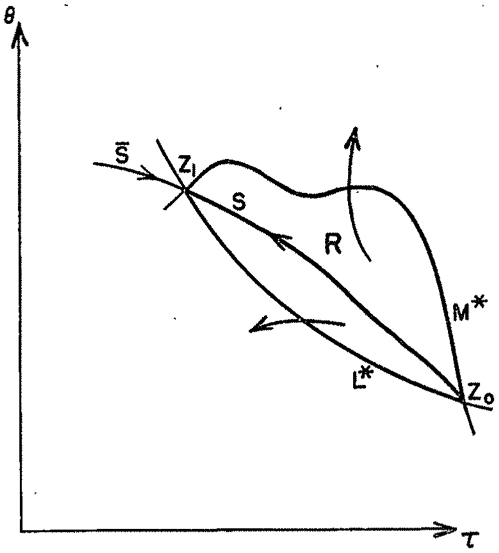


FIG. 1.

represented by the set of "equivalent" shock layers, $S(x + h)$, $(h = \text{constant})$. A shock layer will be called *parametrized* if a particular representative of this class is designated, and, unless so noted, equivalent shock layers will be considered identical.

We observe that if an integral of (E) has the limits Z'_0, Z'_1 at $x = -\infty$, $+\infty$, respectively, then Z'_0 Z'_1 must satisfy the shock conditions (4), (5). It is necessary, therefore, to restrict the definition of shock layer, as we have done, to end states defining a normal shock wave.

We consider first the question of existence and uniqueness of the shock layer for arbitrary positive $\lambda = \lambda(\tau, \theta)$, $\mu = \mu(\tau, \theta)$.

By condition (B), Z_0 and Z_1 are the exclusive singular points of the system (E).[4] We prove that Z_0 is an unstable node and Z_1 a saddle point of (E).[5] For, the characteristic equation of the system (E) at Z_0 and Z_1 is

$$(6) \qquad 0 = \begin{vmatrix} L_\theta/\lambda - \kappa & L_\tau/\lambda \\ M_\theta/\mu & M_\tau/\mu - \kappa \end{vmatrix}$$
$$= \kappa^2 - (M_\tau/\mu + L_\theta/\lambda)\kappa + (M_\theta L_\theta/\lambda\mu)(M_\tau/M_\theta - L_\tau/L_\theta),$$

where the values λ, μ, L_τ, M_θ, etc., are to be taken at Z_0 and Z_1. The discriminant of this equation is

$$[(M_\tau/\mu - L_\theta/\lambda)^2 + 4(M_\theta L_\tau/\lambda\mu)]_{Z=Z_0, Z_1};$$

by conditions (A) and (C), M_θ, $L_\tau > 0$, so that the discriminant is positive at both Z_0 and Z_1, and therefore the roots of (6) are real. The constant terms of (6), is, by conditions (A) and (D), positive at Z_0, negative at Z_1. The latter point is therefore a saddle, the former a node, and, since M_τ, $L_\theta > 0$ (conditions (A), (C), (D)), an unstable node.

It follows from well-known general considerations that there are exactly two integral curves of (E) which approach the saddle Z_1 as $x \to +\infty$, and exactly two which approach it as $x \to -\infty$, these pairs corresponding to the negative and positive roots, respectively, of the characteristic equation (6). The two members of each pair have the same slope at Z_1, but approach it from opposite directions. The slopes are given by

$$-L_\tau/(L_\theta - \kappa\lambda) = -(M_\tau - \kappa\mu)/M_\theta,$$

and, in particular, for $\kappa < 0$ at Z_1, this is negative. Hence, by considering the sign of $L(\tau, \theta)/M(\tau, \theta)$ in the neighborhood of Z_1, it is seen that one of the solutions converging to Z_1 as $x \to +\infty$ approaches it from the region R, in which the ratio is negative. Let this solution be designated by $S(x)$. We establish that $S(x)$ is a shock layer. For, consider the integral curves of (E) which pass through the points of M^* and L^*. On M^*, $M(\tau, \theta) = 0$, $L(\tau, \theta) > 0$, so that all integral curves have vertical tangent vectors, and are directed outwards from R for increasing x. Similarly, on L^*, all solutions have zero slope, with $L(\tau, \theta) = 0$, $M(\tau, \theta) < 0$, and are directed outwards from R, since the slope of L^* is negative (by conditions (A) and (C)). Thus, for *decreasing* x, all integral curves of (E) passing through L^* and M^*

---

[4] This makes clear that the shock transition from Z_0 to Z_1 cannot (strictly) occur in a finite x-interval, since the singular points of (E) can be approached only for $x = \pm\infty$.

[5] This is proved by Weyl in the (τ, p) plane ([1], Theorem 6).

are directed *into* R. In R itself, all integral curves have negative slope and are traversed for decreasing x in the direction of increasing τ and decreasing θ. Let us now follow $S(x)$ for decreasing x. By virtue of the preceding, $S(x)$ cannot intersect L^* or M^* between Z_0 and Z_1. Since there are no other singular points of (E), and since $S(x)$ is monotonic while in R, and cannot terminate inside R, it must approach the node Z_0 as $x \rightarrow -\infty$. This proves the existence of a shock layer $S(x) = (\tau(x), \theta(x))$.

To prove the uniqueness of the corresponding shock layer curve S, we show that no other integral curve of (E) can join Z_0 and Z_1. For Z_0, being an unstable node, which can be approached by an integral curve only as $x \rightarrow -\infty$, the only solution other than S which could join Z_0 and Z_1 would have to be the second integral curve, call it $\bar{S}$, which enters Z_1 as $x \rightarrow +\infty$. This is seen to be impossible as follows. If $\bar{S}$ is also a shock layer curve, then the arcs S and $\bar{S}$ form a closed curve bounding a simply connected region G in the Z-plane. One of the two integral curves of (E) which approach Z_1 as $x \rightarrow -\infty$ enters this region. It cannot terminate in G, since Z_0 and Z_1 are the exclusive singular points of (E); nor can it approach asymptotically a limit cycle in G, for this in turn would have to contain in its interior a singular point of (E).[6] The only possibility remaining[6] is that the curve intersect S or $\bar{S}$, thereby contradicting the uniqueness of the integral curves of (E). This contradiction completes the uniqueness proof.

Thus, under assumptions (A)-(D), we can state the following existence and uniqueness theorem.

THEOREM 1. *Let* $Z_0 = (\tau_0, \theta_0)$ *and* $Z_1 = (\tau_1, \theta_1)$ *be states of a fluid which satisfy the shock conditions,* (4), (5); *then, for any viscosity* $\mu = \mu(\tau, \theta)$, *and heat conductivity* $\lambda = \lambda(\tau, \theta)$, *there exists a unique shock layer joining* Z_0 *to* Z_1.

2. Limit behavior of the shock layer as $\lambda \rightarrow 0$, $\mu \rightarrow 0$. For given $\lambda(Z)$, $\mu(Z)$, let the shock layers of the preceding section be designated by $S(x; \lambda, \mu) = (\tau(x; \lambda, \mu), \theta(x; \lambda, \mu))$, and the associated layer curve by $S(\lambda, \mu)$. We assume now concerning $\lambda(Z)$ and $\mu(Z)$ that these coefficients depend on parameters (independent of τ, θ) in such a way that λ and μ can independently be made arbitrarily small in R. In the following, when bounds are placed on λ, μ, or λ/μ, these bounds are to be understood as holding throughout the region R.

---

[6] Kamke, *Differentialgleichungen Reeller Funktionen*, Chelsea, New York (reprinted 1947), pp. 216, 220-222.

It is our chief object in this section to establish that, as λ and μ approach zero (in whatever manner), the corresponding shock layers approximate a shock wave. In a sense, this "justifies" the usual shock wave theory by showing it to be the limit case of the corresponding theory of real fluids. In the details, it is shown that the limits as $\lambda, \mu \to 0$ exist in the sense of function limits. This requires the individual shock layers to be so parametrized—since they are fixed only to within an x-translation—that the appropriate limits exist under this parametrization. We consider first the double limit, under the same assumptions as in Theorem 1.

THEOREM 2. *Let* $Z_0 = (\tau_0, \theta_0)$ *and* $Z_1 = (\tau_1, \theta_1)$ *be initial and final states of the shock wave,* $Z = Z_0$, $-\infty < x < \xi$, $Z = Z_1$, $\xi < x < \infty$. *Then, the corresponding family of shock layers,* $S(x; \lambda, \mu)$, *if suitably parametrized, approaches this shock wave as* $\lambda \to 0$ *and* $\mu \to 0$ *independently; that is,*

$$\begin{aligned}\lim_{\substack{\lambda\to 0\\ \mu\to 0}} S(x;\lambda,\mu) &= Z_0, \quad -\infty < x < \xi,\\ &= Z_1, \quad \xi < x < \infty,\end{aligned}$$

the convergence being uniform in every closed interval not containing $x = \xi$.

Proof.[7] For $\epsilon > 0$, let $R(\epsilon)$ be the subregion of R outside the circles of radius ϵ about Z_0, Z_1. Consider the equations (E) for any integral curve in R in the form,

$$\text{(E)} \qquad d\theta/dx = L(\tau, \theta)/\lambda, \qquad d\tau/dx = M(\tau, \theta)/\mu,$$

and take the difference,

$$d(\theta - \tau)/dx = L(\tau, \theta)/\lambda + |M(\tau, \theta)|/\mu > 0$$

Define $\eta = \text{Max}\,(\lambda, \mu)$ in R, and let $C(\epsilon)$ be a constant for which

$$L(\tau, \theta) + |M(\tau, \theta)| \geq C(\epsilon) > 0, \qquad (\tau, \theta) \,\varepsilon\, R(\epsilon);$$

such a $C(\epsilon)$ exists, since $L(\tau, \theta) > 0$ on M^*, and $|M(\tau, \theta)| > 0$ on L^* (except at Z_0, Z_1), and are also positive in $R(\epsilon)$. Then, we have for any integral curve in $R(\epsilon)$,

$$(7) \qquad d(\theta - \tau)/dx > C(\epsilon)/\eta > 0$$

Now consider any shock layer $S(x; \lambda, \mu)$, and let (τ_M, θ_M) and (τ_m, θ_m) designate the value of $S(x; \lambda, \mu)$ where it intersects the ϵ circles about Z_1

---

[7] The author is indebted to E. Hopf for this proof, which strengthens and simplifies the author's original proof.

and Z_0 respectively, with $S(x_0; \lambda, \mu) = (\tau_m, \theta_m)$ and $S(x_1; \lambda, \mu) = (\tau_M, \theta_M)$. We have, for $x_0 \leq x \leq x_1$,

$$\tau_0 > \tau_m \geq \tau(x; \lambda, \mu) \geq \tau_M > \tau_1;$$
$$\theta_0 < \theta_m \leq \theta(x; \lambda, \mu) \leq \theta_M < \theta_1.$$

From inequality (7) for $S(x; \lambda, \mu)$, it follows,

$$(8) \qquad x_1 - x_0 \leq \eta/C(\epsilon)[\theta_M - \tau_M - (\theta_m - \tau_m)] < \eta/C(\epsilon)[\theta_1 - \theta_0 + \tau_0 - \tau_1],$$

Thus, given ϵ and δ, and $S(x; \lambda, \mu)$ for which

$$\eta = \operatorname{Max}(\lambda, \mu) \leq \delta C(\epsilon)/(\theta_1 - \theta_0 + \tau_0 - \tau_1),$$

the simultaneous inequalities,

$$(9) \qquad |\tau(x; \lambda, \mu) - \tau_{0,1}| > \epsilon, \qquad |\theta(x; \lambda, \mu) - \theta_{0,1}| > \epsilon,$$

can hold only when the values of $S(x; \lambda, \mu)$ lie in $R(\epsilon)$, and therefore in an interval of length less that $x_1 - x_0 < \delta$. In other words, by making heat conductivity and viscosity sufficiently small, the width of the shock front can be made arbitrarily small. To complete the proof of the theorem, one need only observe that the solutions $S(x; \lambda, \mu)$ can be so parametrized that the intervals (x_0, x_1) contain $x = \xi$. The convergence of the $S(x; \lambda, \mu)$ to the discontinuous shock is then uniform outside of every open interval containing $x = \xi$.

We note that with arbitrary parametrization, either the double limit is a shock wave, as above, or does not exist in an entire interval, or is a constant, Z_0 or Z_1. Hence, among all parametrizations, the only non-trivial limits of shock layers, as $\lambda, \mu \to 0$ independently, are the shock waves.

We note also that (8) gives a (probably very poor) upper bound on the width of the shock transition in the sense of inequality (9). Similarly, lower bounds as well can be obtained from reverse inequalities of the type (7). It is possible that, by taking other linear combinations of the two equations (E), and thereby obtaining inequalities analogous to (7), more satisfactory bounds can be achieved.

We turn now to the single limit problems, $\lim_{\lambda \to 0} S(x; \lambda, \mu)$ and $\lim_{\mu \to 0} S(x; \lambda, \mu)$, for fixed $\mu(Z)$ and $\lambda(Z)$, respectively. First we prove,

LEMMA 1. *Let G be any open neighborhood of the closed arc L^*. Then for sufficiently small values of λ/μ, all shock layer curves $S(\lambda, \mu)$ lie entirely in G.*

Proof. Let $\bar{L}$ be an arc of bounded negative slope with endpoints $\neq Z_0$, Z_1 on M^*, and lying so close to L^* that the subregion of R contained between L^* and $\bar{L}$ contains only points of G. Such a curve $\bar{L}$ can always be found by virtue of the monotonicity of L^* between Z_0 and Z_1. Let D be the subregion of R bounded by $\bar{L}$ and M^*; D contains the complement of G in R. On $\bar{L}$, $L(\tau, \theta) \geq k_1 > 0$, $|M(\tau, \theta)| \leq k_2$, and $|\text{slope}\, \bar{L}| \leq N$, where k_1, k_2, and N are suitable positive constants. Let now $\mu/\lambda > (k_2/k_1)N$; then since the slopes of the corresponding integral curves of (E) satisfy on $\bar{L}$ the inequality,

$$-d\theta/d\tau = -(\mu/\lambda)(L(\tau, \theta)/M(\tau, \theta)) \geq (\mu/\lambda)(k_1/k_2) > N,$$

we have that these integral curves must be directed into D. Thus, any integral curve of (E) for which $\mu/\lambda > (k_2/k_1)N$ and which contains a point of D cannot intersect $\bar{L}$ beyond this point (i. e. for increasing x), and hence cannot pass through Z_1. As a consequence, if $S(\lambda, \mu)$ is a shock layer curve for which $\lambda/\mu < k_1/k_2N$, $S(\lambda, \mu)$ must lie entirely in the region G (indeed, in the smaller region between L^* and $\bar{L}$).

Similar considerations apply in case of small μ/λ. An interesting difference arises from the possibility that M^* is not monotonic. If the arc is monotonic, then exactly the same argument as above applies, with but evident verbal changes. If, however, M^* is not monotonic between Z_0 and Z_1, consider then the arc $\bar{M}^*$ defined by the monotonic function in $[\tau_1, \tau_0]$,

$$\theta = \bar{m}(\tau) = \operatorname*{Min}_{\tau_1 \leq t \leq \tau} m(t); \qquad \tau_1 \leq \tau \leq \tau_0.$$

This arc (Fig. 2) joining Z_1 and Z_0[8] encloses between it and L^* a subregion of R in which, we assert, all shock layer curves must be contained. For, if $\theta = \sigma(\tau)$ is the equation of any shock layer, we have, for every τ in $[\tau_1, \tau_0]$, and some t such that $\tau_1 \leq t \leq \tau$,

$$\sigma(\tau) \leq \sigma(t) \leq m(t) = \bar{m}(\tau),$$

with the inequality, $\sigma(\tau) < \bar{m}(\tau)$, holding provided $\tau \neq \tau_0, \tau_1$. Hence, all shock layer curves lie below $\bar{M}^*$ in R. (The same proof shows that $\bar{M}^*$ is the maximum among all monotonic arcs dominated by M^*.) Replacing M^* by $\bar{M}^*$, we find that the same argument as in the case that M^* is monotonic proves that, for sufficiently small μ/λ, the shock layer curves $S(\lambda, \mu)$ lie in any preassigned neighborhood of $\bar{M}^*$. Thus we can state,

---

[8] That $Z_0 \varepsilon \bar{M}^*$ is clear; for, if not, there is a $\bar{\tau}$, $\tau_1 < \bar{\tau} < \tau_0$, for which $l(\bar{\tau}) < m(\bar{\tau}) < m(\tau_0) = l(\tau_0)$, contradicting the monotonic decreasing character of $l(\tau)$.

LEMMA 2. *Let $\bar{M}^*$ be the arc defined by the function,*

$$\theta = \bar{m}(\tau) = \underset{\tau_1 \leq t \leq \tau}{\mathrm{Min}}\ m(t), \qquad (\tau_1 \leq \tau \leq \tau_0);$$

and let G be any open neighborhood of $\bar{M}^*$. Then all shock layer curves $S(\lambda, \mu)$, for which μ/λ is sufficiently small, lie entirely in G.

Consider now the reduced systems,

$$\text{(10a)} \quad \begin{aligned} \mu(d\tau/dx) &= M(\tau, \theta) \\ 0 &= L(\tau, \theta) \end{aligned}\ ; \qquad \text{(10b)} \quad \begin{aligned} \lambda(d\theta/dx) &= L(\tau, \theta) \\ 0 &= M(\tau, \theta). \end{aligned}$$

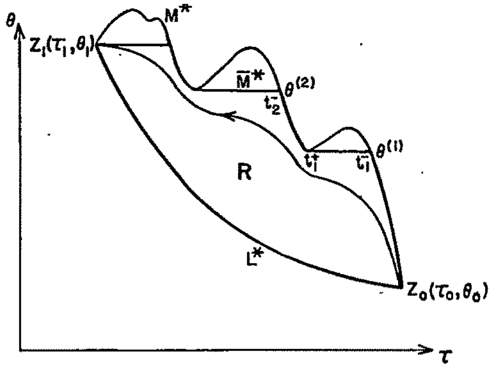


FIG. 2.

The former, (10a), are the equations of a viscous, thermally non-conducting flow, and (10b) of a non-viscous, thermally conducting flow. Let us call any solution of the reduced system (10a), for τ in $[\tau_1, \tau_0]$ and for all x, a *thermally non-conducting shock layer*; (the constant solutions, $(\tau(x), \theta(x)) = Z_0, Z_1$, are thus included in this definition, chiefly for convenience). Then we have,

THEOREM 3. *Let $\bar{S}(x; \bar{\mu})$ be a thermally non-conducting shock layer with viscosity $\bar{\mu}$. Then, the family of shock layers, $S(x; \lambda, \bar{\mu})$, if suitably parametrized, approaches $\bar{S}(x; \bar{\mu})$ as $\lambda \to 0$; that is,*

$$\lim_{\lambda \to 0} S(x; \lambda, \bar{\mu}) = \bar{S}(x; \bar{\mu}),$$

the convergence being uniform in x in every finite interval.

Proof. Let $\zeta = \bar{S}(\bar{x}; \bar{\mu}) = (\bar{\tau}(\bar{x}; \bar{\mu}), \bar{\theta}(\bar{x}; \bar{\mu}))$ be a point on L^*, and x_λ, values such that $x_\lambda \to \bar{x}$ as $\lambda \to 0$. Choose the parametrization of the shock layers, $S(x; \lambda, \bar{\mu}) = (\tau(x; \lambda, \bar{\mu}), \theta(x; \lambda, \bar{\mu}))$, so that $\lim_{\lambda \to 0} S(x_\lambda; \lambda, \bar{\mu}) = \zeta$; this is possible by virtue of Lemma 1. Also from this lemma, we have for every shock layer,

$$\theta = l(\tau) + \epsilon(\tau, \lambda), \tag{11}$$

where $\lim_{\lambda \to 0} \epsilon(\tau, \lambda) = 0$ uniformly in $\tau_1 \leq \tau \leq \tau_0$, and therefore in $-\infty < x < +\infty$. Thus,

$$\bar{\mu} d\tau(x; \lambda, \bar{\mu})/dx = M(\tau, l(\tau) + \epsilon(\tau, \lambda)).$$

Because of the choice of the $\tau(x; \lambda, \bar{\mu})$, we have from this equation, that $\lim_{\lambda \to 0} \tau(x; \lambda, \bar{\mu})$ exists uniformly in every finite x-interval and satisfies

$$\bar{\mu} d\tau/dx = M(\tau, l(\tau)),$$

that is, satisfies the reduced system (10a). Since there is a unique solution of the reduced system passing through the point ζ for $x = \bar{x}$, we conclude $\lim_{\lambda \to 0} \tau(x; \lambda, \bar{\mu}) = \bar{\tau}(x; \bar{\mu})$, and from (11) $\lim_{\lambda \to 0} \theta(x; \lambda, \bar{\mu}) = \bar{\theta}(x; \bar{\mu})$, these limits, by the above, being uniform in every finite interval.

We observe from the above proof that if the shock layers are so parametrized that $\lim_{\lambda \to 0} S(x; \lambda, \bar{\mu})$ exists for a single value of x, then it exists for all x and defines a thermally non-conducting shock layer.

For λ fixed and $\mu \to 0$ the limit solution of the reduced system is no longer continuous in general, as the following Theorem 4 shows. This points up a basic difference in the effects of viscosity and heat conduction on the structure of the shock layers.

To simplify considerations, it will be assumed in the following that M^* has only a finite number of minima, so that the arc $\bar{M}^*$ contains at most a finite number of intervals on which θ is constant. We note that Z_0 cannot lie in such an interval, while Z_1 may or may not, in general. In particular, for an ideal gas, M^* is a parabola; if it is not monotonic, $\bar{M}^*$ consists of the segment of $\theta = \theta_1$ intercepted by M^* plus the arc of M^* joining the segment to Z_0.

Let the function $\theta = \bar{m}(\tau)$, $(\tau_1 \leq \tau \leq \tau_0)$, be constant on the intervals, $[t_i^+, t_i^-]$, $(i = 1, \cdots, n)$, which are ordered so that $t_i^- > t_i^+ > t_{i+1}^-$, (Fig. 2). We enlarge the notion of solution of the reduced system (10b) by admitting certain *discontinuous* solutions, namely, functions $\bar{\tau}(x)$, $\bar{\theta}(x)$ satisfying (10b) for $\bar{\tau}$ in $[\tau_1, \tau_0]$, except at points $x = x_i$, $i = 1, \cdots, n$, where $\bar{\tau}(x_i - 0) = t_i^-$, $\bar{\tau}(x_i + 0) = t_i^+$; these solutions are uniquely determined in $-\infty < x < +\infty$ up to an x-translation. If $n = 0$, that is, if $m(t)$ is strictly monotonic, then the solutions are, of course, continuous. Any solution of (10b) for τ in

$[\tau_1, \tau_0]$ and for all x, whether continuous, or discontinuous in the above sense, we term a *non-viscous shock layer.* We now prove,

THEOREM 4. *Let $\bar{S}(x; \bar{\lambda})$ be a non-viscous shock layer with heat conductivity $\bar{\lambda}$. Then, the family of shock layers, $S(x; \bar{\lambda}, \mu)$, if suitably parametrized, approaches $\bar{S}(x; \bar{\lambda})$ as $\mu \to 0$; that is,*

$$\lim_{\mu \to 0} S(x; \bar{\lambda}, \mu) = \bar{S}(x; \bar{\lambda}),$$

the convergence being uniform in x in any closed interval not containing a point of discontinuity of $\bar{S}(x; \bar{\lambda})$.

Proof. If M^* is strictly monotonic, the theorem is proved exactly as Theorem 3. If, however, this is not the case, we proceed as follows. Let $\bar{S}(x; \bar{\lambda}) \equiv \bar{S}(x) = (\bar{\tau}(x), \bar{\theta}(x))$; (for notational convenience in the following, we omit reference to $\bar{\lambda}$, since it will be the common value of the heat conductivity for all the shock layers under consideration). Let $\zeta = \bar{\tau}(\bar{x}), \bar{\theta}(\bar{x}))$ be a point on M^*, $(\bar{x} < \bar{x}_1)$, where $\bar{x}_1$ is the first point of discontinuity of $\bar{\tau}(x)$. For any set x_μ, such that $x_\mu \to \bar{x}$ as $\mu \to 0$, let the shock layers $S(x; \mu) = (\tau(x; \mu), \theta(x; \mu))$ be so parametrized that $\lim_{\mu \to 0} S(x_\mu; \mu) = \zeta$. We have, by Lemma 2,

$$\tau = m^{-1}(\theta) + \epsilon(\theta, \mu), \tag{12}$$

where $m^{-1}(\theta)$ is the inverse of $m(\tau)$ for $t_1^- \leq \tau \leq \tau_0$, $(\theta_0 \leq \theta \leq \theta^{(1)} = m(t_1^-))$, and $\lim_{\mu \to 0} \epsilon(\theta, \mu) = 0$ uniformly in every closed interval of $\theta_0 \leq \theta < \theta^{(1)}$. For $\theta(x; \mu)$ we therefore have,

$$\bar{\lambda} d\theta(x; \mu)/dx = L[m^{-1}(\theta) + \epsilon(\theta, \mu), \theta]. \tag{13}$$

Since $\theta(x_\mu, \mu) \to \theta(\bar{x})$ and $x_\mu \to \bar{x}$ as $\mu \to 0$, and $\bar{\tau} = m^{-1}(\bar{\theta})$ for $\theta_0 \leq \theta < \theta^{(1)}$, if follows that $\lim_{\mu \to 0} \theta(x; \mu) = \bar{\theta}(x)$, $-\infty < x < \bar{x}_1$, and from (12), $\lim_{\mu \to 0} \tau(x; \mu) = \bar{\tau}(x)$, the convergence in both cases being uniform in every closed interval to the left of $x = \bar{x}_1$. To prove convergence for $x > \bar{x}_1$, consider first the case that $\tau_1 \neq t_1^+$. By the preceding there are values $x_\mu^{(1)} < \bar{x}_1$, such that $x_\mu^{(1)} \to \bar{x}_1$, $\tau(x_\mu^{(1)}, \mu) \to t_1^-$, and $\theta(x_\mu^{(1)}, \mu \to \theta^{(1)}$, as $\mu \to 0$. With the same parametrization of the $S(x; \mu)$ as chosen above, let ξ_μ be values for which $\theta(\xi_\mu; \mu) = \theta^{(1)}$; then $\tau(\xi_\mu; \mu) \to t_1^+$ as $\mu \to 0$ (Lemma 2). We show now that $| \xi_\mu - x_\mu^{(1)} | \to 0$ with μ. For, $L(\tau, \theta)$ is bounded away from zero in the neighborhood of the segment $\theta = \theta^{(1)}$ of $\bar{M}^*$; say $L(\tau, \theta) > k > 0$ in such a neighborhood; and, if μ is sufficiently small, the arcs of $S(x; \mu)$ for

$x_\mu^{(1)} \leq x \leq \xi_\mu$ lie in this neighborhood (Lemma 2). For these μ, we have from the first member of (E),

$$(14) \quad |\,\xi_\mu - x_\mu^{(1)}\,| \leq (\lambda/k)[\theta(\xi_\mu, \mu) - \theta(x_\mu^{(1)}, \mu)] = (\lambda/k)[\theta^{(1)} - \theta(x_\mu^{(1)}, \mu)],$$

which approaches zero as $\mu \to 0$. Hence $\xi_\mu \to \bar{x}_1$. We consider equation (12) and (13) again, where now $m^{-1}(\theta)$ is the inverse of $\theta = m(\tau)$ for $t_2^- \leq \tau \leq t_1^+$, $(\theta^{(1)} \leq \theta \leq \theta^{(2)} = m(t_2^-))$, and $\lim\limits_{\mu\to 0} \epsilon(\theta, \mu) = 0$ uniformly in every closed interval of $\theta^{(1)} \leq \theta < \theta^{(2)}$. As before, since $\theta(\xi_\mu, \mu) \to \theta^{(1)} = \bar{\theta}(x_1)$ as $\xi_\mu \to \bar{x}_1$, and $\bar{\tau} = m^{-1}(\bar{\theta})$ for $\theta^{(1)} \leq \bar{\theta} \leq \theta^{(2)}$, and also since the solution of

$$\lambda d\theta/dx = L(m^{-1}(\theta), \theta)$$

is unique for $\theta(\bar{x}_1) = \theta^{(1)}$, and $\theta^{(1)} \leq \theta \leq \theta^{(2)}$, it follows that $\lim\limits_{\mu\to 0} \theta(x; \mu) = \bar{\theta}(x)$ and $\lim\limits_{\mu\to 0} \tau(x; \mu) = \bar{\tau}(x)$ for $\bar{x} \leq x < \bar{x}_2$, the convergence being uniform in the closed intervals of $(\bar{x}_1, \bar{x}_2)$. One proceeds in this way until the arc of $\bar{M}^*$ containing Z_1 is reached. If Z_1 is not contained in one of the intervals of $\bar{M}^*$, then the above process shows the uniform convergence of $S(x; \mu)$ to $(\bar{\tau}(x), \bar{\theta}(x))$ in every closed subinterval of the half line $\bar{x}_n < x < \infty$. If, however, Z_1 belongs to one of the intervals of $\bar{M}^*$, namely $\theta = \theta_1$, then, by definition of the discontinuous non-viscous shock layer, $\tau_1 = t_n^+$, and $(\bar{\tau}(x), \bar{\theta}(x)) = (\tau_1, \theta_1) = Z_1$ for all x in $\bar{x}_n < x < \infty$. In this case, we prove the convergence, $S(x; \mu) \to Z_1$ for $\bar{x}_n < x < \infty$, as follows. For any small $\delta > 0$, let $R(\delta)$ be the subregion of R outside the circle of radius δ about Z_1. Let $k(\delta)$ be such that $L(\tau, \theta) \geq k(\delta) > 0$ in a neighborhood of $\theta = \theta_1$ in $R(\delta)$. If now the values $x_\mu^{(n)} < \bar{x}_n$ are such that $x_\mu^{(n)} \to \bar{x}_n$, $\tau(x_\mu^{(n)}, \mu) \to t_n^-$, and $\theta(x_\mu^{(n)}, \mu) \to \theta^{(n)} = \theta_1$, as $\mu \to 0$, then, for any arguments $\xi_\mu > x_\mu^{(n)}$ such that $S(\xi_\mu; \mu) \in R(\delta)$, it follows from the first of equations (E),

$$|\,\xi_\mu - x_\mu^{(n)}\,| \leqq (\lambda/k(\delta))[\theta_1 - \theta(x_\mu^{(n)}; \mu] \to 0, \text{ as } \mu \to 0;$$

in other words, those $x > \bar{x}_n$ for which $\tau(x; \mu) - \tau_1 > \delta$ lie in an interval about $\bar{x}_n$ which grows arbitrarily small as $\mu \to 0$. Since δ is arbitrary, this proves the uniform convergence, $\tau(x; \mu) \to \tau_1$, $\theta(x; \mu) \to \theta_1$, in any closed half line of $\bar{x}_n < x < \infty$, and completes the proof of the theorem.

One observes from the above proof that if the shock layers are so parametrized that $\lim\limits_{\mu\to 0} S(x; \lambda, \mu)$ exists for a single value of x, then it exists for all x and defines a non-viscous shock layer.

To complete discussion of the limit behavior of the shock layers, it remains to establish the existence of the iterated limits as $\lambda, \mu \rightarrow 0$, and to show their equality with the double limit (and therefore with a shock wave). We require an appropriate parametrization of the shock layers, with respect to which the double limit and the two iterated limits exist simultaneously. For this, take a circle of sufficiently small radius ϵ about Z_0 (it suffices $\epsilon < \tau_0 - t_1^-$); this intersects each shock layer curve, $S(\lambda, \mu)$, at exactly one point, which we designate by $\zeta_{\lambda\mu}$; and let ζ_λ, ζ_μ represent the point of intersection with M^* and L^*, respectively. We note that $\lim\limits_{\lambda\rightarrow 0} \zeta_{\lambda\mu} = \zeta_\mu$ (Lemma 1), and $\lim\limits_{\mu\rightarrow 0} \zeta_{\lambda\mu} = \zeta_\lambda$ (Lemma 2). For fixed ξ, now let

$$x_{\lambda\mu} = \xi - \tfrac{1}{2}\delta_{\lambda\mu}, \qquad \bar{x}_\mu = \xi - \tfrac{1}{2}\delta_\mu, \qquad \bar{x}_\lambda = \xi - \tfrac{1}{2}\delta_\lambda,$$

where,

$$\delta_{\lambda\mu} = (\eta/C(\epsilon))[\theta_1 - \theta_1 + \tau_0 - \tau_1], \qquad \eta = \mathrm{Max}(\lambda, \mu),$$

$$\delta_\mu = (\mathrm{Max}\, \mu/C(\epsilon))[\theta_1 - \theta_0 + \tau_0 - \tau_1],$$

$$\delta_\lambda = (\mathrm{Max}\, \lambda/C(\epsilon))[\theta_1 - \theta_0 + \tau_0 - \tau_1],$$

$C(\epsilon)$ being the same bound as in inequality (7) of Theorem 2. We make the assignment $S(x_{\lambda\mu}; \lambda, \mu) = \zeta_{\lambda\mu}$; this will prove to be the desired parametrization. From the proof of Theorem 2, we see that, since ξ is contained in all the intervals $(x_{\lambda\mu}, x_{\lambda\mu} + \delta_{\lambda\mu})$ as $\eta \rightarrow 0$, it must follow,

$$\begin{aligned} \lim_{\substack{\lambda\rightarrow 0\\ \mu\rightarrow 0}} S(x; \lambda, \mu) &= Z_0, \quad -\infty < x < \xi \\ &= Z_1, \quad \xi < x < +\infty. \end{aligned}$$

Also, since for fixed μ,

$$\lim_{\lambda\rightarrow 0} \delta_{\lambda\mu} = (\mathrm{Max}\, \mu/C(\epsilon))[\theta_1 - \theta_0 + \tau_0 - \tau_1] = \delta_\mu,$$

we have $\lim\limits_{\lambda\rightarrow 0} x_{\lambda\mu} = \xi - \frac{1}{2}\delta_\mu = \bar{x}_\mu$, $\lim\limits_{\lambda\rightarrow 0} S(x_{\lambda\mu}; \lambda, \mu) = \zeta_\mu$. Therefore as in the proof of Theorem 3, $\lim\limits_{\lambda\rightarrow 0} S(x; \lambda, \mu) = (\bar{\tau}(x; \mu), \theta(x; \mu))$ where $(\bar{\tau}(x; \mu), \bar{\theta}(x; \mu))$ is the solution of the reduced system (10a) which satisfies $(\bar{\tau}(\bar{x}_\mu; \mu), \bar{\theta}(\bar{x}_\mu; \mu)) = \zeta_\mu$. This, combined with the existence of the double limit, establishes the existence of the iterated limit, $\lim\limits_{\mu\rightarrow 0} (\lim\limits_{\lambda\rightarrow 0} S(x; \lambda, \mu))$, and its equality with the double limit. Identical considerations apply to $\lim\limits_{\lambda\rightarrow 0} (\lim\limits_{\mu\rightarrow 0} S(x; \lambda, \mu)$. Consequently, we may state, THEOREM 5. *Let* $Z_0 = (\tau_0, \theta_0)$, $Z_1 = (\tau_1, \theta_1)$ *be initial and final states of the shock wave,* $Z = Z_0$, $-\infty < x < \xi$, $Z = Z_1$, $\xi < x < +\infty$. *Then the shock layers* $S(x; \lambda, \mu)$ *can be parametrized so that,*

$$\lim_{\substack{\lambda\to 0\\ \mu\to 0}} S(x;\lambda,\mu) = \lim_{\lambda\to 0} (\lim_{\mu\to 0} S(x;\lambda,\mu))$$

$$= \lim_{\mu\to 0} (\lim_{\lambda\to 0} S(x;\lambda,\mu)) = \begin{array}{l} Z_0, \ -\infty < x < \xi, \\ Z_1, \ \xi < x < +\infty. \end{array}$$

It is clear, conversely, that the iterated limit, if it exists (for some parametrization), must be either a constant or a shock wave.

In the case of polytropic gases with λ, μ taken independent of τ and θ, the convergence of the iterated limits to a shock wave is visible from explicit formulas for the thermally non-conducting shock layers, $\lim_{\lambda\to 0} S(x;\lambda,\mu)$, and for the non-viscous shock layers, $\lim_{\mu\to 0} S(x;\lambda,\mu)$, (presupposing, however, the existence of these single limits; see Becker [2], for example).

The advantages of the (τ, θ) plane as phase plane should be noted. In these variables, the right members of equations (E), and therefore the curves L and M, are independent of λ and μ. This fact makes the proof of Theorem 1 quite simple, and also permits study, with respect to a fixed geometry, of the dependence of the shock layers on λ and μ. In contrast, the topologies of other phase planes, such as the (τ, p) plane used by Weyl in [1], are such as to obscure a satisfactory proof of the existence and limit theorems.

3. Proof of Conditions (A) — (D). It remains to show that conditions (A)-(D) of Section 1 are satisfied for the fluids under consideration. We examine first the polytropic gases because of their particular importance and simplicity. For polytropic gases, $e(\tau, \theta) = \alpha\theta$, $p(\tau, \theta) = \beta\theta/\tau$, where $\alpha, \beta > 0$ are constants of the medium. Thus

$$L(\tau,\theta) = b[\alpha\theta - b^2/2(\tau - a)^2 - c], \quad M(\tau,\theta) = 1/b[\beta\theta/\tau + b^2(\tau - a)].$$

The Z-plane is the quadrant $\tau > 0$, $\theta > 0$, and the constants a, b, c are assumed to be such that the parabolas L: $L(\tau, \theta) = 0$, and M: $M(\tau, \theta) = 0$, intersect in two points, $Z_0 = (\tau_0, \theta_0)$, $Z_1 = (\tau_1, \theta_1)$, $\tau_0 > \tau_1$. Conditions (A)-(D) are now easily seen to hold. (A) $M_\theta, L_\theta > 0$, since $\alpha, \beta, b > 0$; (B) the parabolas L, M have only one component in the Z-plane, and cannot intersect in more than the two assumed points, Z_0, Z_1; (C) $L_\tau = -b^3/2(\tau - a) > 0$ for $\tau < a$, hence for $\tau_1 \leq \tau \leq \tau_0 < a$; (D) follows from the fact that M lies above L between their points of intersection.

We turn now to the general class of fluids defined by Weyl. These are characterized by the following conditions:

I. $d\tau/dp\big)_{S=\text{const.}} < 0$, $S = \text{entropy}$;

Ia. $S_p(\tau, p) > 0$, $\theta_p(\tau, p) > 0$;

II. $d^2\tau/dp^2\big)_{S=\text{const.}} > 0$;

III. in the continuous process of adiabatic compression one can raise pressure arbitrarily high;

IV. the thermodynamic state Z is uniquely specified by pressure p and specific volume τ, and the points (τ, p) representing the possible states Z in a (τ, p) diagram form a convex region.

To these conditions must of course be added the basic thermodynamic relation,

V. $de = \theta dS - p d\tau$.

This class of fluids evidently includes the polytropic gases, and, as Weyl has shown, the theory of shock waves generalizes in entirety to these fluids.

For fluids subject to the above conditions, Weyl proves a number of important results, among which we need the following.

(a) The Hugoniot contour,

$$H(Z, Z_0) = e(\tau, p) - e(\tau_0, p_0) - \tfrac{1}{2}(p + p_0)(\tau_0 - \tau) = 0$$

is a simple curve on which $s = (p - p_0)/(\tau_0 - \tau)$ grows strictly monotonically from $m_0 = -dp/d\tau\big)_{S=S_0}$ at Z_0 to $+\infty$ as Z moves on the upper branch, $p > p_0$, $\tau < \tau_0$.

(b) If $H(Z_1, Z_0) = 0$, Z_1 lying on the upper branch of $H(Z, Z_0) = 0$, then the straight line $(p - p_0)/(\tau_0 - \tau) = (p_1 - p_0)/(\tau_0 - \tau_1) = b^2$ intersects the adiabatic $S = S_0$ at a point $Z = Z_A$, Z_1 lying between Z_0 and Z_A, and $m_0 < b^2 < m_1 = -dp/d\tau(Z_1)\big|_{S=S_1}$.

To prove conditions (A)-(D) of Section **1**, consider first the mapping $(\tau, p) \rightarrow (\tau, \theta)$ defined by $\theta = \theta(\tau, p)$, $\tau = \tau$. This mapping is 1-1 as well as continuous, for, if both (τ_1, p_1) and (τ_2, p_2) map into the same point (τ, θ), then $\tau_1 = \tau = \tau_2$, and, since $\theta_p > 0$ in the convex (τ, p) plane, it follows also that $p_1 = p_2$. Consequently, the mapping is topological and the $Z(\tau, \theta)$ plane is simply connected.

In the following, let $Z_0(\tau_0, p_0)$, $Z_1(\tau_1, p_1)$, be points for which $H(Z_1, Z_0) = 0$, with $p_1 > p_0$, $\tau_1 < \tau_0$; (either $p_1 > p_0$, $\tau_1 < \tau_0$, or $p_1 < p_0$, $\tau_1 > \tau_0$). Let

$$bM(\tau, p) \equiv p + b^2(\tau - a), \qquad (1/b)L(\tau, p) \equiv e(\tau, p) - \tfrac{1}{2}b^2(\tau - a)^2 - c,$$

$b > 0$, a, c being determined by relations (4), (5). The functions $M(\tau, \theta)$, $L(\tau, \theta)$ are defined in the obvious way:

$$M(\tau, \theta) \equiv M(\tau, p(\tau, \theta)), \qquad L(\tau, \theta) \equiv L(\tau, p(\tau, \theta)).$$

Condition (A) is easily proved, for

$$bM_\theta(\tau, \theta) = p_\theta(\tau, \theta) > 0, \qquad \text{(by Ia)},$$

and

$$\begin{aligned}(1/b)L_\theta(\tau, \theta) = e_\theta(\tau, \theta) &= \theta S_\theta(\tau, \theta), \qquad \text{(by V)}\\ &= S_p(\tau, p)p_\theta(\tau, \theta) > 0,\end{aligned}$$

so that $L_\theta(\tau, \theta) > 0$.

To prove (B), we observe from (a) that $Z_0(\tau_0, p_0)$, $Z_1(\tau_1, p_1)$ are the only simultaneous solutions of $L(\tau, p) = 0$, $M(\tau, p) = 0$ (Weyl's Theorem 7), and therefore $Z_0(\tau_0, \theta_0)$, $Z_1(\tau_1, \theta_1)$ are the only solutions of $L(\tau, \theta) = 0$, $M(\tau, \theta) = 0$. Also, $M(\tau, p) = 0$ only on the straight line M joining Z_0 and Z_1, and since the (τ, p) plane is convex, M consists of only one component. The set in the (τ, θ) plane on which $M(\tau, \theta) = 0$ must therefore also consist of a single curve, namely, the image of M. Letting $\mathscr{L}$ designate the set of points (τ, p) on which $L(\tau, p) = 0$, it remains only to show that the component $L \subset \mathscr{L}$, which passes through Z_1 also passes through Z_0. Since

$$(1/b)L_p(\tau, p) = e_p(\tau, p) = \theta S_p(\tau, p) > 0,$$

the curve L is representable as a single valued function $p = l(\tau)$, with slope,

$$\begin{aligned}(15) \qquad dl(\tau)/d\tau &= -(e_\tau(\tau, p) - b^2(\tau - a))/e_p(\tau, p)\\ &= -(\theta S_\tau(\tau, b) - p - b^2(\tau - a))/\theta S_p(\tau, p), \quad ((\tau, p) \varepsilon L).\end{aligned}$$

In particular, at Z_1, L has the slope,

$$-S_\tau/S_p(\tau_1, p_1) = dp/d\tau(Z_1)\Big|_{S=S_1} = -m_1 < -b^2.$$

Consequently, L, for increasing τ, enters the convex region D contained between the adiabatic $S = S_0$ and the straight line M (Fig. 3). Since D

lies below M, $p + b^2(\tau - a) < 0$ in D, so that, by (15), $l(\tau)$ is monotonically decreasing function of τ in this region. The curve L cannot terminate in D, must therefore intersect either M, which can only occur at Z_0, or the adiabatic, $S = S_0$.

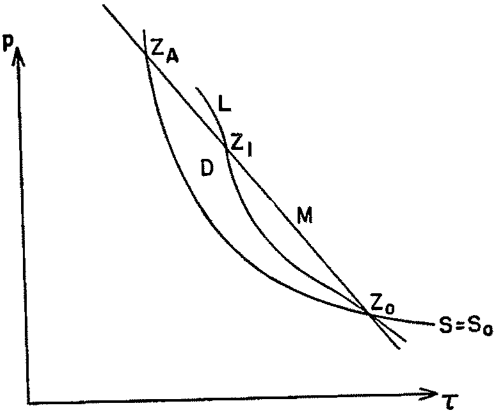


FIG. 3.

But on $S = S_0$,

$$(1/b)\,(dL(\tau,p)/d\tau)\Big)_{S=S_0} = de/d\tau\Big)_{S=S_0} - b^2(\tau - a)$$

$$= [-p - b^2(\tau - a)]_{S=S_0} > 0, \text{ for } \tau_A < \tau < \tau_0;$$

hence $L(\tau, p) \neq 0$ on $S = S_0$ between Z_0 and Z_1. L must therefore pass through Z_0 as well as Z_1. This proves that condition (B) is satisfied.

For condition (C),

$$(1/b)L_\tau(\tau,\theta) = e_\tau(\tau,\theta) - b^2(\tau - a) = \theta S_\tau(\tau,\theta) - p - b^2(\tau - a);$$

but from V, we have, $S_\tau(\tau,\theta) = p_\theta(\tau,\theta) > 0$,

$$\therefore\ (1/b)L_\tau(\tau,\theta) = \theta p_\theta(\tau,\theta) - p - b^2(\tau - a) > 0 \text{ on } L, \text{ for } \tau_1 \leqq \tau \leqq \tau_0.$$

(D) follows from the relations,

$$M_\tau/M_\theta(\tau,\theta) - L_\tau/L_\theta(\tau,\theta) = [M_\tau/M_p(\tau,p) - L_\tau/L_p(\tau,p)]\theta_p(\tau,p)$$

$$= (b^2 - m_0)\theta_p(Z_0) > 0 \quad \text{at } Z_0$$

$$= (b^2 - m_1)\theta_p(Z_1) < 0 \quad \text{at } Z_1.$$

This completes the proof that a Weyl fluid satisfies the conditions (A)-(D) which were assumed in obtaining Theorems 1-5.

BIBLIOGRAPHY.

[1] H. Weyl, "Shock waves in arbitrary fluids," *Communications on Pure and Applied Applied Mathematics*, vol. 2 (1949), pp. 103-122.

[2] R. Becker, "Stosswelle und Detonation," *Zeitschrift für Physik*, vol. 8 (1921-22), pp. 321-362.

[3] L. H. Thomas, "Note on Becker's theory of the shock front," *Journal of Chemical Physics*, vol. 12 (1944), pp. 449-452.

[4] Lord Rayleigh, "Aerial plane waves of finite amplitude," *Proceedings of the Royal Society of London*, Ser. A 84 (1910), pp. 247-284.

[5] A. E. Puckett and H. J. Stewart, "The thickness of a shock wave in air," *Quarterly of Applied Mathematics*, vol. 7 (1950), pp. 457-463.

[6] M. Morduchow and P. Libby, "On a complete solution of the one-dimensional flow equations of a viscous, heat-conducting, compressible gas," *Journal of the Aeronautical Sciences*, vol. 16 (1949), pp. 674-684, 704.

[7] R. Courant and K. Friedrichs, *Supersonic flow and shock waves*, Interscience, New York (1948).

[8] G. I. Taylor and J. W. Maccoll, *The Mechanics of compressible fluids*, vol. III of W. F. Durand, *Aerodynamic Theory*, Springer, Berlin (1934), pp. 218-222.

[9] P. Duhem, *Recherches sur l'hydrodynamique*, Gauthier-Villars, Paris (1903); Première partie, Ch. I. = Annales de la Faculté des Sciences Toulouse, (2) vol. 3 (1901), pp. 315-377.

EQUIVALENCES OF MEASURE SPACES.*

By I. E. Segal.

1. Introduction. For any measure space we define the "measure ring" as the Boolean ring of all measurable sets (including not necessarily σ-finite sets) modulo the ideal of all null sets, the "σ-finite measure ring" as the subring of elements which are residue-classes of σ-finite sets, and the "multiplication algebra" as the algebra of all multiplications of elements of the Hilbert space of square integrable functions over the space by bounded measurable functions. Two measure spaces have algebraically isomorphic measure rings (we say then that they are "strongly equivalent") if and only if their multiplication algebras are unitarily equivalent, or alternatively, if their multiplication algebras are algebraically isomorphic. This is of special interest in view of the fact [5] that every maximal abelian self-adjoint algebra of bounded operators on a Hilbert space is unitarily equivalent to the multiplication algebra of some measure space. The multiplication algebra of a measure space is maximal abelian if and only if its measure ring is complete (as a partially ordered set), or alternatively, if and only if it is strongly equivalent to a direct sum (see below) of finite measure spaces. The class of measure spaces with these properties (we call such spaces "localizable") constitutes in some ways a more natural generalization of the σ-finite measure spaces, than the class of arbitrary measure spaces. In particular, for a measure space to be localizable is equivalent to the validity for the space of the conclusion of the Radon-Nikodym theorem, or alternatively to the validity of the conclusion of the Riesz representation theorem for continuous linear functionals on the Banach space of integrable functions. Every measure space is metrically equivalent (by which we mean that there is a measure-preserving isomorphism between the σ-finite measure rings—roughly speaking this means the spaces are equivalent as far as integration over them is concerned) to a localizable space, and this latter space is essentially unique.

More specifically, every measure space is metrically equivalent to a unique perfect measure space, this being a regular locally compact space in which for every measurable set of finite measure there is a unique continuous function on the space equal nearly everywhere to the characteristic

---

* Received June 25, 1950.

function of the set; and any metric equivalence between localizable spaces is uniquely extendable to a strong equivalence between them. These perfect measure spaces, which are a development of spaces treated by Stone, Freudenthal, and especially Kakutani [3] (the metric equivalence of a finite measure space with a perfect space being proved here), are useful in eliminating certain types of measurability problems (cf. [4]) for they have many of the properties of discrete spaces.

The present paper arose from an investigation of the decomposition and multiplicity theory of algebras of operators on Hilbert spaces, and as the results are of some independent interest they are being published separately. Among other side results we mention the fact that any uniformly locally compact regular measure space, in particular any locally compact group under Haar measure, is localizable. Any two measure spaces have algebraically isomorphic σ-finite measure rings if and only if the corresponding Lebesgue spaces over them are isomorphic as partially ordered linear spaces, and are metrically equivalent if and only if there is a more restricted type of isomorphism between certain function spaces over the measure spaces.

2. Technical preliminaries. In this section we fix the notation and terminology for the remainder of the paper and prove some simple results which are basic in the following.

DEFINITIONS 2.1. A *conditional σ-ring* $\mathcal{R}$ of subsets of a set R is a collection of subsets which is closed under relative complementation and union, and has the further property that if $\{E_n\}$ is a sequence of elements in R such that $E_n \subset E$ for all n, where $E \varepsilon R$, then $\bigcup_n E_n \varepsilon \mathcal{R}$. A *measure space* $(R, \mathcal{R}\ r)$ is the system composed of a set R, a conditional σ-ring $\mathcal{R}$ of subsets of R, and a countably-additive, finite (non-negative) measure r defined on $\mathcal{R}$, such that if $\{E_n\}$ is a sequence of mutually disjoint sets in $\mathcal{R}$ for which the series $\Sigma_n r(E_n)$ is convergent, then $\bigcup_n E_n \varepsilon \mathcal{R}$.

Remark 2.1. The term "measure space" is commonly used in different senses, e. g. frequently in reference to a system $(S, \mathcal{S}, s)$ composed of a set S, a σ-ring $\mathcal{S}$ of subsets of S (i. e. a conditional σ-ring which is closed under arbitrary countable unions), and a not-necessarily-finite, countably additive measure s on $\mathcal{S}$. Any such measure space becomes a measure space in the sense of Definition 2.1 when its σ-ring $\mathcal{S}$ of sets is replaced by the subring $\mathcal{S}_0$ of sets in $\mathcal{S}$ of finite measure, and this replacement affects neither the (suitably-defined) integrability of functions on the space, nor the values of the integrals of integrable functions. It is advantageous, for formal and notational purposes, to replace $\mathcal{S}$ by $\mathcal{S}_0$ because the elements of $\mathcal{S}$-$\mathcal{S}_0$ must generally be treated in terms of the elements of $\mathcal{S}_0$. We mention also that to replace $\mathcal{S}$ by the σ-ring of σ-finite elements of $\mathcal{S}$ (i. e. elements which are countable unions of elements of $\mathcal{S}_0$) is less advantageous, mainly because the distinction between σ-finite and arbitrary elements of $\mathcal{S}$-$\mathcal{S}_0$ is often an artificial one, from the viewpoint of the present paper.

DEFINITIONS 2. 2. *Sets in measure spaces.* A set K in a measure space $M = (R, \mathcal{R}, r)$ (i. e. a subset of R) is called *measurable* if $K \cap E \varepsilon \mathcal{R}$ for all $E \varepsilon \mathcal{R}$; a *null set* if $K \cap E$ is an element of $\mathcal{R}$ on which r vanishes, for all $E \varepsilon \mathcal{R}$. The *measure of a measurable set* K in M is defined as the least upper bound (finite or ∞) of $r(E)$, as E varies over the subsets of K which are in $\mathcal{R}$; and is denoted by $r(K)$. A measurable set of finite measure is called a *chunk.* A statement which is a function on a measure space is said to be true *nearly everywhere* (abbreviated "n. e.") if it is valid at all points of the space, except for a null set. Two subsets K and L of a measure space are said to be *equivalent* if their symmetric difference $(K\text{-}L) \cup (L\text{-}K)$, which will be denoted as $K \ominus L$, is a null set. A set K in M is called *σ-finite* if it is contained in a countable union of chunks. M is called *σ-finite* if R is *σ-finite* and *finite* in case R is itself a chunk.

THEOREM 2. 1. *The collection of all measurable subsets of a measure space is a complemented σ-ring on which the measure is countably additive.*

The proof of this uses standard methods in a straightforward fashion, and so we omit details.

DEFINITIONS 2. 3. *Functions on measure spaces.* A complex-valued function on a measure space is said to be *measurable* if it is the pointwise limit of a sequence of finite linear combinations of characteristic functions of measurable sets. Such a function is said to *vanish at infinity* on the space if for every positive number e, the set where the function exceeds e in absolute value is contained in a chunk. A complex valued function f on a measure space $M = (R, \mathcal{R}, r)$ is said to be *integrable* if there exists a sequence f_n of finite linear combinations of characteristic functions of elements of $\mathcal{R}$ which converges pointwise to f and which is such that $\int | f_n(p) - f_m(p)| \, dr(p)$ converges to 0 as $m, n \to \infty$, where for any finite linear combination $h = \Sigma \alpha_i \chi_{E_i}$ of characteristic functions of mutually disjoint elements E_i of $\mathcal{R}$ (for any set S, χ_S denotes the characteristic function of S), $\int h(p) \, dr(p)$ is defined as $\Sigma \alpha_i r(E_i)$; the value of $\lim_n \int f_n(p) \, dr(p)$ is then called the value of the integral of f, and denoted as $\int f(x) \, dr(x)$. A complex-valued function f on

M is said to be α-th-power integrable (or square-integrable, in case $\alpha = 2$), where $\alpha \geq 1$, if it is measurable and $|f(x)|^{\alpha}$ is integrable. For any such function the value of $\{\int |f(x)|^{\alpha} dr(x)\}^{1/\alpha}$ is called the *norm* of f in L_{α}, and denoted by $\| f \|_{\alpha}$; the collection of all such functions is designated as $L_{\alpha}(M)$. The Banach space consisting of the quotient of the normed linear space $L_{\alpha}(M)$ by the subspace of functions in $L_{\alpha}(M)$ of zero norm is called the (Banach) space of α-th-power integrable functions on M, and designated $B_{\alpha}(M)$. A complex-valued function f on M is said to be *essentially bounded* if there is a real number α such that $|f(x)| \leq \alpha$ except on a null set, and the g. l. b. of all such α is called the *essential bound* of f, denoted as $\| f \|_{\infty}$, the collection of all such functions which are measurable being denoted as $L_{\infty}(M)$. The Banach space, and algebra, consisting of the quotient of the normed space and algebra $L_{\infty}(M)$ (multiplication being the usual pointwise product), by the subspace, and ideal consisting of elements of zero norm, is called the (Banach) *algebra of bounded measurable functions on M*, and denoted $B_{\infty}(M)$. Two functions on a measure space are called *equivalent* if their values coincide n. e. The (real) subspace of all real-valued functions in $B_{\alpha}(M)$, partially-ordered by the definition: $f \leq g$ if $f(x) \leq g(x)$ n. e., is called the *partially-ordered (Banach) space of α-th-power integrable functions on M*, and denoted $RB_{\alpha}(M)$.

DEFINITIONS 2.4. *Measure rings.* A *measure ring* $(\mathfrak{M}, m)$ is the system composed of a σ-Boolean ring $\mathfrak{M}$ (i. e. a Boolean ring in which each countable subset has a l. u. b.) and a countably additive measure m on $\mathfrak{M}$, such that only the zero element has measure zero, and every element is the l. u. b. of elements of finite measure. An element of a measure ring is called *σ-finite* if it is the union (l. u. b.) of a sequence of elements of $\mathfrak{M}$ of finite measure, and a measure ring is called *σ-finite* if each of its elements is.

The *measure ring* $(\mathfrak{M}, m)$ of a measure space $M = (R, \mathfrak{R}, r)$ is defined to have as $\mathfrak{M}$ the quotient of the ring of measurable sets in M by the ideal of null sets, with the measure of any residue class defined to be the measure of any representative of the class. The natural mapping of the ring of measurable sets in M onto $\mathfrak{M}$ will be denoted by θ, and the residue class corresponding to a measurable set will also be denoted by superposing a bar: thus $\theta(K) = \bar{K}$ for any measurable set K.

Remark 2.2. It is easy to see that the mapping θ just defined is a σ-homomorphism, i. e. for any sequence of measurable sets $\{E_n; n = 1, 2, \cdots\}$, $\theta(\bigcup_n E_n) = \bigcup_n \theta(E_n)$ and $\theta(\bigcap_n E_n) = \bigcap_n \theta(E_n)$. It follows that the measure ring of a measure space is a σ-Boolean ring. It will later be shown that every element in $\mathfrak{M}$ is the l. u. b. of elements of finite measure, so that the measure ring of M is actually a measure ring in the sense of our earlier definition.

DEFINITIONS 2. 5. *Measure rings and equivalences.* The *σ-finite (or finite, resp.) measure ring* of a measure space is the subring of the measure ring of the space consisting of all σ-finite elements (or elements of finite measure). Two measure spaces are (*strongly* or *weakly,* resp.) equivalent if there is an algebraic isomorphism between the (Boolean rings of) their (measure rings, or σ-finite measure rings resp.); and *metrically equivalent* if there is a measure-preserving algebraic isomorphism between their finite measure rings. An isomorphism which implements such an equivalence between measure spaces is called a *strong, weak, or metric equivalence,* depending on its character. A measure space $M = (R, \mathfrak{R}, r)$ is said to be *isomorphic* to a measure space $M' = (R', \mathfrak{R}', r')$ if there is a one-to-one mapping ϕ from R onto R' such that: $E \varepsilon \mathfrak{R}$ if and only if $\phi(E) \varepsilon \mathfrak{R}'$; and for $E \varepsilon \mathfrak{R}$, $r(E) = r'(\phi(E))$. Such a mapping ϕ is called an *isomorphism* on M to M'.

DEFINITION 2. 6. A measure space is called *localizable* if its measure ring is complete (as a partially-ordered set).

DEFINITION 2. 7. *Regular and perfect spaces.* A *regular* (*locally compact*) *measure space* is the system composed of a locally compact Hausdorff space R, and a measure space $(R, \mathfrak{R}, r)$ with the properties that: a) $\mathfrak{R}$ contains all compact subsets of R (so that r is finite on such sets); b) for any set E in $\mathfrak{R}$ there is an open set G in $\mathfrak{R}$ containing E and $r(E) = \text{g. l. b.}[E \subset H, H \text{ open}]r(H)$ (we frequently denote the set over which a bound is taken by placing it between the symbol for the bound and the function whose bound is indicated); c) every set in R is in the σ-ring generated by the compact subsets of R. Such a space is called *perfect* if a) for any chunk F there is a continuous function on R which vanishes at infinity on R in the topological sense (i. e. for every positive number e, the set where the function is not less than e in absolute value is compact), and is equivalent to the characteristic function of F; b) the measure of any non-void open set in $\mathfrak{R}$ is positive.

Remark 2. 2. It is easy to see that condition b) of the immediately preceding definition is equivalent to the condition that the continuous function described in condition a) be unique.

Examples 2. 1. As examples of localizable spaces we mention a) σ-finite measure spaces, b) a locally compact group relative to Haar measure (see

Corollary 5. 2. 1). To obtain a perfect space, let R have the discrete topology, let $\mathfrak{R}$ consist of all finite subsets of R, and let $r(E)$ be the number of points in E. An example of a space which can be seen without difficulty to be non-localizable is given in [2], p. 131, Note 9.

We conclude this section by showing that either strong or metric equivalence is stronger, actually, than weak equivalence. The following definition is useful in this and other connections.

DEFINITION 2. 8. A Boolean ring is called *countably-decomposable* if any family of mutually disjoint elements of the ring is at most countable. An element of a Boolean ring is called *countably-decomposable* if the principal ideal which it generates is such.

THEOREM 2. 2. *An element of a measure ring is σ-finite if and only if it is countably-decomposable.*

To show that E is countably decomposable if it is a σ-finite element of the measure ring $(\mathfrak{M}, m)$ of the measure space M, assume the contrary. Then $\bar{E} = \bigcup_i \bar{E}_i$, where E_i is a chunk $(i = 1, 2, \cdots)$ and there exists a collection $\{F_\xi; \xi \in \Xi\}$ of measurable sets in M such that $0 < \bar{F}_\xi \leq \bar{E}$, $\bar{F}_\xi \cap \bar{F}_{\xi'} = 0$ if $\xi \neq \xi'$, and with the index set Ξ more than countable. It follows that $\bar{F}_\xi = \bigcup_i \bar{F}_\xi \cap \bar{E}_i$, for any ξ, and hence that the collection of all $\bar{F}_\xi \cap \bar{E}_i$, as ξ and i vary, is more than countable. It is easy to conclude that for some positive integer n, and some integer i_0, the set $[\xi \mid m(\bar{F}_\xi \cap \bar{E}_{i_0}) > n^{-1}]$ is more than countable. This contradicts the circumstance that the $\bar{F}_\xi \cap \bar{E}_{i_0}$ are mutually disjoint and bounded by the element $\bar{E}_{i_0}$ of finite measure.

Now if $\bar{E}$ is a countably-decomposable element of $\mathfrak{M}$, let $\{\bar{F}_\xi; \xi \in \Xi\}$ be a maximal family of mutually disjoint elements of $\mathfrak{M}$ of finite measure which are not greater than $\bar{E}$; such a family exists by the Zorn principle, and Ξ is at most countable. The maximality of this family implies that $\bar{E} - \bigcup_\xi \bar{F}_\xi$ is not greater than any element of $\mathfrak{M}$ of finite measure, from which it follows that $\bar{E} = \bigcup_\xi \bar{F}_\xi$.

COROLLARY 2. 2. 1. *The contraction of a strong equivalence to the ideal of σ-finite elements is a weak equivalence. In particular, strongly equivalent spaces are weakly equivalent.*

Since countable decomposability is defined in purely algebraic terms, an algebraic isomorphism maps countably-decomposable elements into countably-decomposable elements. By the theorem just proved, this implies that the contraction of a strong equivalence maps the σ-finite elements onto the σ-finite elements.

THEOREM 2.3. *Let* $\mathfrak{F}$ *and* $\mathfrak{F}'$ *be the finite measure rings, and* $\mathfrak{S}$ *and* $\mathfrak{S}'$ *the* σ*-finite measure rings, of the measure spaces* M *and* M' *respectively. Any ring-isomorphism of* $\mathfrak{F}$ *onto* $\mathfrak{F}'$ *can be uniquely extended to a ring-isomorphism of* $\mathfrak{S}$ *onto* $\mathfrak{S}'$.

We put η for a ring isomorphism of $\mathfrak{F}$ onto $\mathfrak{F}'$, and define a mapping ψ on $\mathfrak{S}$ to $\mathfrak{S}'$ as follows: if $\bar{E} \varepsilon \mathfrak{S}$ and $\bar{E} = \bigcup_i \bar{E}_i$, where $\bar{E}_i$ $(i = 1, 2, \cdots)$ is in $\mathfrak{F}$, then $\psi(\bar{E}) = \bigcup_i \eta(\bar{E}_i)$. Clearly $\psi(\bar{E})$ is defined for every $\bar{E}$ in $\mathfrak{S}$, and we now show that ψ coincides on $\mathfrak{F}$ with η. Let $\bar{E}$ be arbitrary in $\mathfrak{F}$, and suppose that $\bar{E} = \bigcup_i \bar{F}_i$, where $\bar{F}_i$ $(i = 1, 2, \cdots)$ is in $\bar{\mathfrak{F}}$. Then $\bar{E} \geqq \bar{F}_i$, whence $\eta(\bar{E}) \geqq \eta(\bar{F}_i)$, and $\eta(\bar{E}) \geqq \bigcup_i \eta(\bar{F}_i) = \psi(\bar{E})$. Setting $\bar{G}' = \eta(\bar{E}) - \psi(\bar{E})$, then $\bar{G}' \cap \eta(\bar{E}) = \bar{G}'$ and $\bar{G}' \cap \eta(\bar{F}_i) = 0 \quad (i = 1, 2, \cdots)$. Now $\bar{G}' = \eta(\bar{G})$, for some $\bar{G}$ in $\mathfrak{F}$, and it results from the equations involving $\bar{G}'$ that $\bar{G} \cap \bar{E} = \bar{G}$ and $\bar{G} \cap \bar{F}_i = 0$ for all i, whence $\bar{G} = \bar{G} \cap \bigcup_i \bar{F}_i = 0$.

Next we show that ψ is single-valued. Let $\bar{E}_i$ and $\bar{F}_i$ $(i = 1, 2, \cdots)$ be in $\mathfrak{F}$, and suppose that $\bigcup_i \bar{E}_i = \bigcup_i \bar{F}_i = \bar{D}$. Let $\bigcup_i \eta(\bar{E}_i) = \bar{A}'$ and $\bigcup_i \eta(\bar{F}_i) = \bar{B}'$. To show that $\bar{A}' = \bar{B}'$, it suffices to show that $\bar{A}' \cap \bar{H}' = \bar{B}' \cap \bar{H}'$ for an arbitrary $\bar{H}'$ in $\mathfrak{F}'$. If $\bar{H}$ is arbitrary in $\mathfrak{F}$, then $\eta(\bar{H})$ is arbitrary in $\mathfrak{F}'$, and hence we need only show that $\eta(\bar{H}) \cap \bigcup_i \eta(\bar{E}_i) = \eta(\bar{H}) \cap \bigcup_i \eta(\bar{F}_i)$. Now

$$\begin{aligned} \eta(\bar{H}) \cap \bigcup_i \eta(\bar{E}_i) &= \bigcup_i \eta(\bar{H}) \cap \eta(\bar{E}_i) = \bigcup_i \eta(\bar{H} \cap \bar{E}_i) \\ &= \psi(\bigcup_i \bar{H} \cap \bar{E}_i) = \psi(\bar{H} \cap \bar{D}), \end{aligned}$$

which equals $\eta(\bar{H} \cap \bar{D})$ since $\bar{H} \cap \bar{D} \varepsilon \mathfrak{F}$. As the last expression is independent of the representation of $\bar{D}$ as a countable union of elements of $\mathfrak{F}$, it follows that ψ is single-valued.

To show ψ to be a ring homomorphism, it suffices to show that $\psi(\bar{A} \cap \bar{B}) = \psi(\bar{A}) \cap \psi(\bar{B})$ and $\psi(\bar{A} \cap \bar{B}) = \psi(\bar{A}) \cap \psi(\bar{B})$, for arbitrary $\bar{A}$ and $\bar{B}$ in $\mathfrak{S}$, say $\bar{A} = \bigcup_i \bar{A}_i$ and $\bar{B} = \bigcup_i \bar{B}_i$, where $\bar{A}_i$ and $\bar{B}_i$ $(i = 1, 2, \cdots)$ are in $\mathfrak{F}$. Now

$$\begin{aligned} \psi(\bar{A} \cap \bar{B}) = \psi(\bigcup_{i,j} \bar{A}_i \cap \bar{B}_j) &= \bigcup_{i,j} \eta(\bar{A}_i \cap \bar{B}_j) = \bigcup_{i,j} \eta(\bar{A}_i) \cap \eta(\bar{B}_j) \\ &= \{\bigcup_i \eta(\bar{A}_i)\} \cap \{\bigcup_j \eta(\bar{B}_j)\} = \psi(\bar{A}) \cap \psi(\bar{B}). \end{aligned}$$

In a similar fashion it follows that $\psi(\bar{A} \cap \bar{B}) = \psi(\bar{A}) \cup \psi(\bar{B})$. To see that ψ is onto $\mathfrak{S}'$, let E' be arbitrary in $\mathfrak{S}'$, and let $\bar{E}' = \bigcup_i \bar{E}_i'$, with $\bar{E}_i' \varepsilon \mathfrak{F}'$. It is easy to conclude that $\bar{E}' = \psi(\bar{E})$, where $\bar{E} = \bigcup_i \eta^{-1}(\bar{E}_i)$. To complete the proof that ψ is an isomorphism, it remains only to show that ψ is univalent; this is easily seen, for if $\psi(\bar{A}) = \psi(\bar{B})$ with $\bar{A}$ and $\bar{B}$ in $\mathfrak{S}$, then $\psi(\bar{A} \ominus \bar{B}) = 0$, whence $\bar{A} \ominus \bar{B} = 0$, or $\bar{A} = \bar{B}$.

Finally ψ is unique, for let ψ_1 be any ring isomorphism of $\mathcal{S}$ onto $\mathcal{S}'$ which extends η. If $\bar{E}$ is arbitrary in $\mathcal{S}$, say $\bar{E} = \bigcup_i \bar{E}_i$, with $\bar{E}_i \varepsilon \mathfrak{F}$ $(i = 1, 2, \cdots)$, then $\bar{E} \geqq \bar{E}_i$, $\psi_1(\bar{E}) \geqq \psi_1(\bar{E}_i) = \eta(\bar{E}_i)$, whence

$$\psi_1(\bar{E}) \geqq \bigcup_i \eta(\bar{E}_i) = \psi(\bar{E}).$$

By the same argument applied to ψ_1^{-1}, ψ^{-1}, and η^{-1} we have $\psi_1^{-1}(\bar{E}') \geqq \psi^{-1}(\bar{E}')$, for $\bar{E}' \varepsilon \mathcal{S}'$. Putting $\bar{E}' = \psi(\bar{E})$, it results that $(\psi_1^{-1}\psi)(\bar{E}) \geqq \bar{E}$, which implies on applying ψ_1 to both sides of the inequality, that $\psi(\bar{E}) \geqq \psi_1(\bar{E})$. Combining the two inequalities shows that $\psi = \psi_1$.

COROLLARY 2. 8. 1. *Metrically equivalent measure spaces are likewise weakly equivalent.*

3. Direct sums of measure spaces. In order to treat measure spaces which are not necessarily σ-finite, it is frequently convenient to represent such spaces in terms of finite spaces. This section is concerned mainly with representations of a measure space as a "direct sum" of other measure spaces, especially of finite ones, where "direct sum" is defined as follows.

DEFINITION 3. 1. Let $M_\xi = (R_\xi, \mathcal{R}_\xi, r_\xi)$ be a family of measure spaces depending on an index ξ which ranges over a set Ξ, and suppose that R_ξ and $R_{\xi'}$ are disjoint when $\xi \neq \xi'$. The *direct sum* $M = (R, \mathcal{R}, r)$ of the M_ξ, as ξ ranges over Ξ, is the measure space for which $R = \bigcup_\xi R_\xi$; $\mathcal{R}$ is the collection of all subsets E of R such that $E \cap R_\xi = 0$ except for an at most countable set Ξ_0 of ξ's, and for which $\sum_{\xi \varepsilon \Xi_0} r_\xi(E \cap R_\xi)$ is convergent; and r is given by the equation $r(E) = \sum_{\xi \varepsilon \Xi_0} r_\xi(E \cap R_\xi)$. The direct sum is denoted as $\Sigma_\xi \oplus M_\xi$.

THEOREM 3. 1. *If the measure space $M_\xi = (R_\xi, \mathcal{R}_\xi, r_\xi)$ is weakly, or strongly, or metrically, respectively equivalent to the measure space $M_\xi' = (R_\xi', \mathcal{R}_\xi', r_\xi')$, for all ξ in Ξ, where $R_{\xi_1} \cap R_{\xi_2} = 0 = R_{\xi_1}' \cap R_{\xi_2}'$ for any two distinct elements ξ_1 and ξ_2 of Ξ, then the direct sum of the M_ξ $(\xi \varepsilon \Xi)$ is weakly, or strongly, or metrically equivalent to the direct sum of the M_ξ' $(\xi \varepsilon \Xi)$.*

Consider first the case of weak equivalence. Let $\mathfrak{F}_\xi$ and $\mathfrak{F}_\xi'$ be the σ-finite measure rings of M_ξ and M_ξ' respectively, and $\mathfrak{F}$ and $\mathfrak{F}'$ the σ-finite measure rings of the direct sums $M = (R, \mathcal{R}, r)$ and $M' = (R', \mathcal{R}, r')$ of the M_ξ and the M_ξ' respectively. Let ϕ_ξ be an algebraic isomorphism of $\mathfrak{F}_\xi$ onto $\mathfrak{F}_\xi'$, and let E be an arbitrary σ-finite measurable set in M. Then $\bar{E}$ is an arbitrary element of $\mathfrak{F}$. To show that M and M' are weakly equivalent amounts showing that $\mathfrak{F}$ and $\mathfrak{F}'$ are algebraically isomorphic. We define an isomorphism ϕ from $\mathfrak{F}$ onto $\mathfrak{F}'$ as follows. Set $E = \bigcup_{\xi \epsilon \Xi_1} E_\xi$, where Ξ_1 is countable, and $E_\xi = E \cap R_\xi$; then $\bar{E} = \bigcup_{\xi \epsilon \Xi_1} \bar{E}_\xi$ and we define $\phi(\bar{E}) = \bigcup_{\xi \epsilon \Xi_1} \phi_\xi(\bar{E}_\xi)$. Next we show that ϕ is uniquely defined. In the first place, if $E = \bigcup_{\xi \epsilon \Xi_2} E_\xi$, where Ξ_2 is another countable subset of Ξ, then plainly $\Xi_1 \cap \Xi_2 \supset \Xi_0$, where $\Xi_0 = [\xi \mid \bar{E}_\xi \neq 0]$, and it is trivial that $\bigcup_{\xi \epsilon \Xi_1} \phi_\xi(\bar{E}_\xi) = \bigcup_{\xi \epsilon \Xi_0} \phi_\xi(\bar{E}_\xi) = \bigcup_{\xi \epsilon \Xi_2} \phi_\xi(\bar{E}_\xi)$. Furthermore, if F is a measurable σ-finite set such that $\bar{E} = \bar{F}$, then $\bar{E} \cap \bar{R}_\xi = \bar{F} \cap \bar{R}_\xi$, and putting

$$\Xi' = [\xi \mid \text{either } \bar{E} \cap \bar{R} \neq 0 \text{ or } \bar{F} \cap \bar{R}_\xi \neq 0],$$

it results that

$$\phi(\bar{E}) = \bigcup_{\xi \epsilon \Xi'} \phi_\xi(\bar{E} \cap \bar{R}_\xi) = \bigcup_{\xi \epsilon \Xi'} \phi_\xi(\bar{F} \cap \bar{R}_\xi) = \phi(\bar{F}).$$

Thus ϕ is well-defined, and to conclude the proof that it is an isomorphism it remains only to show that it is order-preserving and onto $\mathfrak{F}'$. It is straightforward to verify that ϕ has these further properties.

The proof is the same in the case of metric equivalence except for the additional feature that ϕ_ξ is then taken to be measure-preserving and that ϕ must then be shown to be likewise measure-preserving. It is easily seen that ϕ is actually of this character.

Finally we treat the case of strong equivalence. Let $\mathfrak{M}_\xi$ and $\mathfrak{M}'_\xi$ be the measure rings of M_ξ and M'_ξ, and $\mathfrak{M}$ and $\mathfrak{M}'$ the measure rings of M and M', respectively. Let ψ_ξ be an algebraic isomorphism of $\mathfrak{M}_\xi$ onto $\mathfrak{M}'_\xi$, and let K be an arbitrary measurable set in M. We define an algebraic isomorphism ψ of $\mathfrak{M}$ onto $\mathfrak{M}'$, by setting $\psi(\bar{K}) = \bigcup_\xi \psi_\xi (\bar{K} \cap \bar{R}_\xi)$; we proceed to show that the indicated least upper bound exists. In fact, if L_ξ' is any measurable set in M_ξ', $\bigcup_\xi \bar{L}_\xi'$ exist and equals $\theta(\bigcup_\xi L_\xi')$. For it is easily verified that $\bigcup_\xi L_\xi'$ is measurable and that $\theta(\bigcup_\xi L_\xi') \geq \bar{L}_\xi'$; and if W' is any measurable set in M' such that $\bar{W}' \geq \bar{L}_\xi'$, then $\bar{W}' \cap \bar{R}_\xi' \geq \bar{L}_\xi' \cap \bar{R}_\xi' = \bar{L}_\xi'$. It follows that $\{\bar{W}' - \theta(\bigcup_\xi L_\xi')\} \cap \bar{R}_\xi' = 0$, whence for any subset Ξ_0 of Ξ which is not more than countable,

$$\{\bar{W}' - \theta(\bigcup_\xi L_\xi')\} \cap \bigcup_{\xi \epsilon \Xi_0} \bar{R}_\xi' = 0 = \{\bar{W}' - \theta(\bigcup_\xi L_\xi')\} \cap \theta(\bigcup_{\xi \epsilon \Xi_0} R_\xi').$$

Since every set E' in $\mathfrak{R}'$ is contained in a set of the form $\bigcup_{\xi \epsilon \Xi_0} R_\xi'$, it results that $\{W' - \theta(\bigcup_\xi L_\xi')\} \cap \bar{E}' = 0$ for all such E', which implies that $W' - \bigcup_\xi L_\xi'$ is a null set, i. e. $\bar{W}' \geq \theta(\bigcup_\xi L_\xi')$. It remains only to show that ψ is onto and order-preserving. That ψ has these properties is easily verified.

Direct sums of finite measure spaces occur frequently (e. g. any σ-finite measure space is such) and have the useful property (localizability) stated in the following theorem. Later, Theorem 5. 1, it will be seen that this property almost characterizes direct sums of finite spaces.

THEOREM 3. 2. *A measure space which is strongly equivalent to a direct sum of finite measure spaces is localizable.*

It is plainly sufficient to show that a direct sum of finite measure spaces is localizable. Consider to begin with the case of a single summand.

LEMMA 3. 2. 1. *A finite measure space is localizable.*

Let $\mathcal{B}$ be any subset of the measure ring $(\mathfrak{M}, m)$ of a finite measure space, and let α be the g. l. b. of the measures of all elements of $\mathfrak{M}$ which are upper bound for $\mathcal{B}$,—by the finiteness of the measure space, such elements exist. Let $\{\bar{E}_i; i = 1, 2, \cdots\}$ be a sequence of upper bounds for $\mathcal{B}$ such that $m(\bar{E}_i) \rightarrow \alpha$. Putting $\bar{F} = \bigcap_{i \leq n} \bar{E}_i$, it is clear that $\bar{F}_n$ is an upper bound for $\mathcal{B}$, and that $\alpha \leq m(\bar{F}_n) \leq m(\bar{E}_n)$. It follows that $m(\bar{F}_n) \rightarrow \alpha$, and hence that $m(\bar{F}) = \alpha$, where $\bar{F} = \bigcap_n \bar{F}_n$, and so $\bar{F}$ is an upper bound for $\mathcal{B}$. Now $\bar{F}$ is the l. u. b. of $\mathcal{B}$, for if $\bar{G}$ is any other upper bound for $\mathcal{B}$, then so is $\bar{F} \cap \bar{G}$, which implies that $m(\bar{F} \cap \bar{G}) = \alpha = m(\bar{F})$, from which it results that $\bar{F} \cap \bar{G} = \bar{F}$ and hence that $\bar{F} \leq \bar{G}$.

The following lemma obviously completes the proof of the theorem.

LEMMA 3. 2. 2. *A direct sum of finite measure spaces is localizable.*

Let M be the direct sum of the measure spaces $M_\xi = (R_\xi, \mathfrak{R}_\xi, r_\xi)$, with ξ ranging over Ξ, and let $\mathcal{B}$ be any collection of elements in the measure ring $\mathfrak{M}$ of M. Let $\mathcal{B}_\xi = [\bar{U} \cap \bar{R}_\xi \mid \bar{U} \in \mathcal{B}]$; then by the preceding lemma $\mathcal{B}_\xi$ has a l. u. b. $\bar{V}_\xi$, where V_ξ can be taken to be a measurable set in M_ξ. Setting $V = \bigcup_\xi V_\xi$, it is easily seen that V is measurable, and as seen in the proof of the preceding theorem, $\bar{V} = \bigcup_\xi \bar{V}_\xi$. Now $\bar{V}$ is a bound for $\mathcal{B}$, for if $\bar{U} \in \mathcal{B}$, then $U = \bigcup_\xi (U \cap R_\xi)$ and again by a fact established in the proof of the preceding theorem, $\bar{U} = \bigcup_\xi (\bar{U} \cap \bar{R}_\xi)$; and as $\bar{V}$ bounds $\mathcal{B}_\xi$, we have in particular, $\bar{V} \geq \bar{U} \cap \bar{R}_\xi$ for all ξ, so that $\bar{V} \geq \bigcup_\xi (\bar{U} \cap \bar{R}_\xi) = \bar{U}$. Moreover, $\bar{V}$ is the l. u. b. of $\mathcal{B}$, for if $\bar{W}$ is any upper bound for $\mathcal{B}$, then it is an upper bound for $\mathcal{B}_\xi$, for each ξ, hence an upper bound for all the $\bar{V}_\xi$, and thus an upper bound for $\bar{V}$.

COROLLARY 3. 2. 1. *A σ-finite measure space is localizable.*

It is clear that a σ-finite measure space is a direct sum of a countable number of finite measure spaces.

The converse of the preceding theorem is included in Theorem 3. 4. The next theorem however is needed in the proof of Theorem 3. 4, and it has some independent interest. It shows that the classification of localizable spaces under strong equivalence is equivalent to their classification under weak equivalence which in turn is reducible (via Theorem 3. 4) to the classification of finite measures spaces under weak equivalence. The latter classification exists in a cogent form in [3a].

THEOREM 3. 3. *Let* $\mathfrak{M}$ *and* $\mathfrak{M}'$ *be the measure rings of localizable measure spaces* M *and* M', *and let* ϕ *be an algebraic isomorphism of the* σ*-finite measure ring of* M *onto that of* M'. *Then* ϕ *can be uniquely extended to an algebraic isomorphism of* $\mathfrak{M}$ *onto* $\mathfrak{M}'$.

LEMMA 3. 3. 1. *Any element* $\bar{K}$ *of the measure ring of a measure space is the least upper bound of the elements of finite measure which it bounds.*

It is clear that $\bar{K}$ is an upper bound for the $\bar{K} \cap \bar{E}$, as E ranges over $\mathfrak{R}$, $(R, \mathfrak{R}, r)$ being the measure space in question. To show that it is least, suppose that L is a measurable set in the measure space such that $\bar{L} \geqq \bar{K} \cap \bar{E}$, for all $E \in \mathfrak{R}$. Then we have $(\bar{K} - \bar{L}) \cap \bar{E} = (\bar{K} \cap \bar{E}) - (\bar{L} \cap \bar{E}) = (\bar{K} \cap \bar{E}) - (\bar{K} \cap \bar{E}) = 0$. It follows that $K - L$ meets every element of $\mathfrak{R}$ in a null set, i. e. is a null set so $\theta(K - L) = 0$. Now $\theta(K - L) = \bar{K} - \bar{L}$, and hence $\bar{K} \leqq \bar{L}$.

To conclude the proof of the theorem, it plainly suffices to establish the following lemma. The lemma has been formulated so as to be applicable to the proof of Theorem 3. 4, and for this reason its statement is slightly more complicated than would be necessary if it required only in connection with the present theorem.

LEMMA 3. 3. 2. *Net* $\mathfrak{N}$ *and* $\mathfrak{N}'$ *be Boolean rings with units and let* $\mathfrak{D}$ *and* $\mathfrak{D}'$ *be ideals in* $\mathfrak{N}$ *and* $\mathfrak{N}'$, *respectively, with the properties that any element of* $\mathfrak{N}$ [*or* $\mathfrak{N}'$] *is the least upper bound of the elements of* $\mathfrak{D}$ [*or* $\mathfrak{D}'$] *which it bounds. Let* ϕ *be an algebraic isomorphism of* $\mathfrak{D}$ *onto* $\mathfrak{D}'$ *and suppose that, as a lattice,* $\mathfrak{N}'$ *is complete. Then* ϕ *can be uniquely extended to an algebraic isomorphism of* $\mathfrak{N}$ *into* $\mathfrak{N}'$, *and if* $\mathfrak{N}$ *is complete, this isomorphism is necessarily onto.*

We define a mapping ψ on $\mathfrak{N}$ to $\mathfrak{N}'$ as follows: for an arbitrary element A of $\mathfrak{N}$, $\psi(A) = \text{l. u. b.}\, [X \in \mathfrak{D}, X \leqq A]\phi(X)$. It is clear that ψ is order-preserving and agrees on $\mathfrak{D}$ with ϕ. We proceed to show that it is univalent (i. e. one-to-one into) and a homomorphism.

To prove that ψ is univalent, it is sufficient to show that if $\psi(A) \geqq \psi(B)$, then $A \geqq B$. Now assuming $\psi(A) \geqq \psi(B)$, if $A \geqq B$ is false, there is a non-zero element C of $\mathfrak{D}$ such that $B - A \geqq C$, since otherwise $(B - A) \cap X = 0$ for all $X \varepsilon \mathfrak{D}$ and from the hypothesis, $B - A = \text{l. u. b.}_{X \varepsilon \mathfrak{D}} (B - A) \cap X = 0$, or $A \geqq B$. Thus $B \geqq C$, $\psi(B) \geqq \psi(C)$, and hence

$$\psi(A) \cap \psi(C) \geqq \psi(B) \cap \psi(C) = \psi(C).$$

On the other hand,

$$\begin{aligned} \psi(A) \cap \psi(C) &= \psi(C) \cap \text{l. u. b.}\, [X \varepsilon \mathfrak{D} \mid X \leqq A] \phi(X) \\ &= \text{l. u. b.}\, [X \varepsilon \mathfrak{D}, X \leqq A] \{\phi(C) \cap \phi(X)\} \\ &= \text{l. u. b.}\, [X \varepsilon \mathfrak{D}, X \leqq A] \phi(C \cap X). \end{aligned}$$

The last expression vanishes because when $X \leqq A$, $C \cap X \leqq C \cap A = 0$. Therefore $\psi(C) = 0$, *i. e.* $\phi(C) = 0$, and since ϕ is an isomorphism, there results the contradiction $C = 0$.

To show that ψ is a homomorphism, let A and B be arbitrary elements of $\mathfrak{N}$, and let D be an arbitrary element of $\mathfrak{D}$. Then

$$\begin{aligned} \psi(D) \cap \psi(A) &= \psi(D) \cap \text{l. u. b.}\, [X \varepsilon \mathfrak{D}, X \leqq A] \phi(X) \\ &= \text{l. u. b.}\, [X \varepsilon \mathfrak{D}, X \leqq A] \phi(D \cap X) = \text{l. u. b.}\, [Y \varepsilon \mathfrak{D}, Y \leqq D \cap A] \phi(Y) \\ &= \psi(D \cap A). \end{aligned}$$

Now

$$\begin{aligned} &\text{l. u. b.}\, [D \varepsilon \mathfrak{D}, D \leqq B] (\psi(D) \cap \psi(A)) \\ &\qquad = \psi(A) \cap \text{l. u. b.}\, [D \varepsilon \mathfrak{D}, D \leqq B] \psi(D) = \psi(A) \cap \psi(B). \end{aligned}$$

On the other hand,

$$\begin{aligned} &\text{l. u. b.}\, [D \varepsilon \mathfrak{D}, D \leqq B] \psi(D \cap A) \\ &\qquad = \text{l. u. b.}\, [X \varepsilon \mathfrak{D}, X \leqq A \cap B] \psi(X) = \psi(A \cap B). \end{aligned}$$

Thus $\psi(A \cap B) = \psi(A) \cap \psi(B)$. A parallel procedure shows that $\psi(A \cup B) = \psi(A) \cup \psi(B)$.

Thus ψ is an algebraic isomorphism of $\mathfrak{N}$ into $\mathfrak{N}'$ and extends ϕ. If ψ_1 is any such isomorphism, and if A is an arbitrary element of $\mathfrak{N}$, the equation $A = \text{l. u. b.}\, [X \varepsilon \mathfrak{D}, X \leqq A] X$ implies that

$$\psi_1(A) \geqq \text{l. u. b.}\, [X \varepsilon \mathfrak{D}, X \leqq A] \phi(X) = \psi(A).$$

In particular, if I and I' are the units of $\mathfrak{N}$ and $\mathfrak{N}'$ respectively, $\psi_1(I) \geqq \psi(I) = I'$, so that $\psi_1(I) = I'$. It follows that $\psi_1(I - A) = I' - \psi_1(A)$, and employing this equation in conjunction with the inequality $\psi_1(I - A)$

$\geq \psi(I - A)$ shows that $\psi(A) \geq \psi_1(A)$. As the opposite semi-inequality has already been obtained, it follows that $\psi = \psi_1$, i. e. ψ has the stated uniqueness property.

Now suppose that $\mathfrak{N}$ is complete. To show that ψ is then onto, let A' be an arbitrary element of $\mathfrak{N}'$; we shall show that there exists an element A of $\mathfrak{N}$ such that $\psi(A) = A'$. Set $A = \text{l. u. b.}\ [X \varepsilon \mathfrak{D}, \phi(X) \leq A']X$. Then $A \geq X$ for $X \varepsilon \mathfrak{D}$ and $\phi(X) \leq A'$, whence $\psi(A) \geq \phi(X)$ for all such X. Since ϕ is onto, A' is the l. u. b. of the $\phi(X)$, for $X \varepsilon \mathfrak{D}$ and $\phi(X) \leq A'$, and hence $\psi(A) \geq A'$. On the other hand, if D' is an element of $\mathfrak{D}'$ such that $D' \leq \psi(A) - A'$, then $D' \leq \psi(A)$, i. e. $D' \leq \text{l. u. b.}\ [X \varepsilon \mathfrak{D}, X \leq A]\phi(X)$, and intersecting this inequality with D' yields the equation

$$D' = \text{l. u. b.}\ [X \varepsilon \mathfrak{D}, X \leq A]\{\phi(X) \cap \phi(D)\},$$

where D is the element of $\mathfrak{D}$ for which $\phi(D) = D'$. Now $\phi(X) \cap \phi(D) = \phi(X \cap D)$, and as X ranges over the elements of $\mathfrak{D}$ which are bounded by A, $X \cap D$ ranges over the elements of $\mathfrak{D}$ which are bounded by $A \cap D$. It results that $D' = \text{l. u. b.}\ [Y \varepsilon \mathfrak{D}, Y \leq A \cap D]\phi(Y)$. Intersecting the equation, $A = \text{l. u. b.}\ [\phi(X) \leq A', X \varepsilon \mathfrak{D}]$, with D yields the equation $A \cap D = \text{l. u. b.}\ [\phi(X) \leq A', X \varepsilon \mathfrak{D}]X \cap D$. If $X \varepsilon \mathfrak{D}$ and $\phi(X) \leq A'$, then since $D' \leq \psi(A) - A'$ we have $D' \cap A' = 0$, and hence $D' \cap \phi(X) = 0 = \phi(X \cap D)$. This implies $X \cap D = 0$, and hence it follows from the equation for $A \cap D$ that $A \cap D = 0$. The equation for D' then shows that $D' = 0$. Thus $\{\psi(A) - A'\} \cap D' = 0$ for all $D' \varepsilon \mathfrak{D}'$. Since any element of $\mathfrak{N}'$ is the l. u. b. of the elements of $\mathfrak{D}'$ which it bounds,

$$\psi(A) - A' = \text{l. u. b.}_{D' \varepsilon \mathfrak{D}'} \{\psi(A) - A'\} \cap D' = 0.$$

In other terms, $A' \geq \psi(A)$, and since the semi-inequality in the opposite direction has been previously obtained, we can conclude that $A' = \psi(A)$.

The following theorem is useful for reducing questions concerning arbitrary measure spaces to corresponding questions for finite measure spaces. In combination with Theorem 3. 2 it shows that localizable spaces are precisely those which are, in some ways, most readily treated in terms of finite spaces, and thereby helps to justify our designation for these spaces (see also condition *e.* of Theorem 5. 1).

THEOREM 3. 4. *For any measure space M there is a measure-preserving algebraic isomorphism of its measure ring $\mathfrak{M}$ into the measure ring $\mathfrak{M}'$ of a direct sum of finite measure spaces, whose contraction to the σ-finite subring of $\mathfrak{M}$ is onto the same subring of $\mathfrak{M}'$. For such an isomorphism to be onto it is necessary and sufficient that M be localizable.*

Let $M = (R, \mathcal{R}, r)$ and let $\{R_\xi; \xi \in \Xi\}$ be a family of chunks in M of positive measure, which is maximal with respect to the property that the intersection of any two sets in the family is of measure zero,—it is clear from Zorn's principle that such a family exists. Let $\mathcal{R}_\xi$ be the subring of $\mathcal{R}$ consisting of the elements of $\mathcal{R}$ which are contained in R_ξ, and let $M_\xi = (R_\xi, \mathcal{R}_\xi, r)$. Let $M_\xi' = (R_\xi', \mathcal{R}_\xi', r_\xi')$ be a space isomorphic to M_ξ whose set R_ξ' is chosen in such a way that $R_{\xi_1}' \cap R_{\xi_2}' = 0$ for $\xi_1 \neq \xi_2$—e. g., R_ξ' may be taken to be the set of all pairs (x, ξ) with $x \in R_\xi$, and with $\mathcal{R}_\xi'$ and r_ξ' determined by the condition that the mapping η_ξ, $x \rightarrow (x, \xi)$, on R_ξ to R_ξ', be an isomorphism on M_ξ to M_ξ'. Now let $M' = (R', \mathcal{R}', r')$ be the direct sum of the M_ξ', $\xi \in \Xi$. It is easily seen that if E is in the family $\mathcal{B}$ of all chunks in M, then the set $\Xi_0 = [\xi \mid r(E \cap R_\xi) \neq 0]$ is at most countable. Since $\sum_{\xi \in \Xi_0} r(E \cap R_\xi) \leq r(E)$, the sum $\sum_{\xi \in \Xi_0} r_\xi'(\eta_\xi(E \cap R_\xi))$ is convergent, and hence $\bigcup_{\xi \in \Xi_0} \eta_\xi(E \cap R_\xi)$, $= E'$ say, is in $\mathcal{R}'$. We define a mapping ϕ_0 on $\mathcal{B}$ to $\mathcal{R}'$ as follows: $\phi_0(E) = E'$, and observe that if E and F are equivalent elements of $\mathcal{B}$, then $\phi_0(E)$ and $\phi_0(F)$ are equivalent elements of $\mathcal{R}'$. It results that the mapping ϕ, on $\mathcal{D} = \theta(\mathcal{B})$ to $\mathcal{D}' = \theta(\mathcal{R}')$, defined by the equation: $\phi(\bar{E}) = \theta(\phi_0(E))$, $E \in \mathcal{B}$, is single-valued. We show next that ϕ is a measure-preserving algebraic isomorphism of $\mathcal{D}$ onto $\mathcal{D}'$.

That ϕ is measure-preserving means that if E is as in the preceding paragraph, then $r(E) = r'(E')$. Now

$$r'(E') = \sum_{\xi \in \Xi_0} r_\xi'(\eta_\xi(E \cap R_\xi)) = \sum_{\xi \in \Xi_0} r(E \cap R_\xi).$$

Putting $G = E - \bigcup_{\xi \in \Xi_0} (E \cap R_\xi)$, it follows that G is a chunk and has measure $r(E) - r'(E')$. It is easy to see that $G \cap R_\xi$ is a null set, for all ξ. The maximality of the collection $\{R_\xi; \xi \in \Xi_0\}$ hence implies that G is a null set, i. e. $r(E) - r'(E') = 0$.

To show that ϕ is a homomorphism, let E_1 and E_2 be arbitrary in $\mathcal{B}$, and let $\Xi_3 = \Xi_1 \cup \Xi_2$, where $\Xi_i = [\xi \mid r(E_i \cap R_\xi) \neq 0]$ $(i = 1, 2)$. Observe that if E is arbitrary in $\mathcal{B}$, as above, and if Ξ_4 is any set such that $\Xi_4 \supset \Xi_3 \cup \Xi_0$ then $\phi(\bar{E}) = \bigcup_{\xi \in \Xi_4} \theta(\eta_\xi(E \cap R_\xi))$, for $E \cap R_\xi$ is a null set when $\xi \in \Xi_4 - \Xi_0$, so that for these values of ξ, $\theta(\eta_\xi(E \cap R_\xi)) = 0$, and hence $\bigcup_{\xi \in \Xi_4} \theta(\eta_\xi(E \cap R_\xi)) = \bigcup_{\xi \in \Xi_0} \theta(\eta_\xi(E \cap R_\xi)) = \theta(\bigcup_{\xi \in \Xi_0} \eta_\xi(E \cap R_\xi)) = \bar{E}'$. It follows that $\phi(\bar{E}_1) \cup \phi(\bar{E}_2) = \{\bigcup_{\xi \in \Xi_3} \theta(\eta_\xi(E_1 \cap R_\xi))\} \cup \bigcup_{\xi \in \Xi_3} \theta(\eta_\xi(E_2 \cap R_\xi))$. Now M' is a direct sum of finite measure spaces, and hence localizable, by Theorem 3. 2,

i. e. its measure ring $\mathfrak{M}'$ is complete. Consequently the right side of the last equation can be expressed in the form

$$\begin{aligned}&\bigcup_{\xi\in\Xi_3}\{\theta(\eta_\xi(E_1\cap R_\xi))\cup\theta(\eta_\xi(E_2\cap R_\xi))\}\\ &\qquad=\bigcup_{\xi\in\Xi_3}\theta(\eta_\xi(E_1\cap R_\xi)\cup\eta_\xi(E_2\cap R_\xi))\\ &\qquad=\bigcup_{\xi\in\Xi_3}\theta(\eta_\xi((E_1\cup E_2)\cap R_\xi)=\phi(\theta(E_1\cup E_2))=\phi(\bar{E}_1\cup\bar{E}_2).\end{aligned}$$

In a similar fashion it can be shown that $\phi(\bar{E}_1)\cap\phi(\bar{E}_2)=\phi(\bar{E}_1\cap\bar{E}_2)$.

It follows that ϕ is an isomorphism, for if $\phi(\bar{E}_1)=\phi(\bar{E}_2)$, where E_1 and E_2 are as above, then $\phi(\bar{E}_1\ominus\bar{E}_2)=0$, so that $\bar{E}_1\ominus\bar{E}_2=0$, or $\bar{E}_1=\bar{E}_2$. Moreover, ϕ is onto $\mathfrak{D}'$, for if F' is in $\mathfrak{R}'$, $F'=\bigcup_{\xi\in\Xi_0}(F'\cap R_\xi')$ for some set Ξ_0 which is at most countable, and setting $F=\bigcup_{\xi\in\Xi_0}\eta_\xi^{-1}(F'\cap R_\xi)$, F is clearly measurable, $r(F)=\sum_{\xi\in\Xi_0}r(\eta_\xi^{-1}(F'\cap R_\xi))=\sum_{\xi\in\Xi_0}r'(F'\cap R'_\xi)=r(F')$ so that F is a chunk, and it is clear that $\phi(\bar{F})=\bar{F}'$.

By Lemma 3. 3. 2 (which is applicable because of Lemma 3. 3. 1), ϕ can be uniquely extended to an algebraic isomorphism ψ of $\mathfrak{M}$ into $\mathfrak{M}'$. Now ϕ is measure-preserving, and ψ agrees with ϕ except on elements of $\mathfrak{M}-\mathfrak{D}$, all of which elements have infinite measure and map into elements of $\mathfrak{M}'-\mathfrak{D}'$, which likewise have infinite measure. Therefore ψ is measure-preserving. It is clear that ψ maps σ-finite elements of $\mathfrak{M}$ into σ-finite elements of $\mathfrak{M}'$ Conversely, if $\bar{E}'$ is any σ-finite element of $\mathfrak{M}'$, then we have $\bar{E}'=\bigcup_i\bar{E}_i'$, where $\{E_i;i=1,2,\cdots\}$ is a sequence of chunks. Now any chunk H in M differs from a countable union of elements in $\mathfrak{R}$ by a set of measure zero, for if $\{H_n;n=1,2,\cdots\}$ is a sequence of elements of $\mathfrak{R}$ such that $r(H_n)$ converges to the least upper bound of the measure of subsets of H which are in $\mathfrak{R}$, then it is easily seen that $H-\bigcup_nH_n$ intersects each element of $\mathfrak{R}$ in a null set and so has measure zero. It follows that E' differs from the union of a sequence $\{F_i';i=1,2,\cdots\}$ of elements of $\mathfrak{R}'$ by a null set. Hence $\bar{E}'=\bigcup_i\bar{F}_i'$ and putting $\bar{E}=\bigcup_i\phi^{-1}(\bar{F}_i)$, it results that $\bar{E}$ is a σ-finite element of $\mathfrak{M}$ such that $\psi(\bar{E})=\bar{E}'$.

Finally, if M is localizable, then $\mathfrak{M}$ is complete,, and by Lemma 3. 3. 2 ψ is necessarily onto. On the other hand, if ψ is onto, then M is strongly equivalent to a direct sum of finite measure spaces, and hence by Theorem 3. 3 is localizable.

4. Equivalences of spaces. In the present section we study the relations between various types of equivalence of measure spaces, and corresponding types of equivalence of function spaces or operator algebras associated with the measure spaces. Before stating our first theorem along these lines we make a number of definitions and introduce some additional notation.

DEFINITIONS AND NOTATIONS 4. 1. *Multiplication algebras.* The *multiplication algebra* of a measure space $M = (R, \mathfrak{R}, r)$ is the algebra of all operators T_k on $B_2(M)$ of the form $(T_k f)(x) = k(x)f(x)$, where f is a general element of $L_2(M)$ and k is an arbitrary bounded measurable function on M. Here, and elsewhere when convenient, we simplify notation by denoting a residue class in $B_a(M)$ by any function in that residue class. T_k is called the *operation of multiplication by k.* Two algebras of bounded linear operators $\mathfrak{A}$ and $\mathfrak{A}'$ on Hilbert spaces $\mathfrak{H}$ and $\mathfrak{H}'$ are called *unitarily equivalent* if there exists a unitary transformation U from $\mathfrak{H}$ onto $\mathfrak{H}'$ such that $U\mathfrak{A} = \mathfrak{A}'U$, and the mapping $W \to UWU^{-1}$ is then called the *algebraic isomorphism of $\mathfrak{A}$ induced by U.* If K is any set in M, χ_K denotes its characteristic function, and if K is measurable, $\chi_{\bar{K}}$ denotes the residue class in $B_\infty(M)$ of which χ_K is an element.

Remark 4. 1. It is obvious that $\chi_{\bar{K}}$ actually depends only on $\bar{K}$, so that our notation is justified. We mention also that it is easy to verify that the mapping $k \to T_k$ on $B_\infty(M)$ to the multiplication algebra of M is an isomorphism.

The following theorem has interesting implications for multiplicity theory, by virtue of the fact that the multiplication algebra of a localizable measure space is essentially the most general maximal abelian SA (self-adjoint) algebra of bounded linear operators on a Hilbert space (cf. [5]). The theorem also includes a kind of analog, for measure spaces, of the Stone-Gelfand-Kolmogoroff theorem associating a homomorphism between compact Hausdorff spaces with an algebraic isomorphism between function spaces over them.

THEOREM 4. 1. *Two measure spaces are strongly equivalent if and only if their algebras of bounded measurable functions are algebraically isomorphic (complex conjugation of functions being taken to be algebraic operation), which in turn is true if and only if their multiplcation algebras are unitarily equivalent.*

More specifically, let M and M' be measure spaces with measure rings $\mathfrak{M}$ and $\mathfrak{M}'$, algebras of bounded measurable functions $\mathfrak{B}$ and $\mathfrak{B}'$, multiplication algebras $\mathfrak{A}$ and $\mathfrak{A}'$. Then if ϕ is a mapping on $\mathfrak{M}$ to $\mathfrak{M}'$ which is a strong equivalence between M and M', there is an algebraic isomorphism $\tilde{\phi}$ from $\mathfrak{B}$ onto $\mathfrak{B}'$ such that $\tilde{\phi}(\chi_K) = \chi_{\phi(K)}$, for all $\bar{K}$ in $\mathfrak{M}$; and conversely every algebraic isomorphism $\tilde{\phi}$ from $\mathfrak{B}$ onto $\mathfrak{B}'$ arises in this fashion.

Furthermore, if $\tilde{\phi}$ *is an algebraic isomorphism from* $\mathscr{B}$ *onto* $\mathscr{B}'$, *then* $\mathcal{A}$ *and* $\mathcal{A}'$ *are unitarily equivalent via a unitary transformation for which the induced algebraic isomorphism of* $\mathcal{A}$ *takes multiplication by* k *into multiplication by* $\tilde{\phi}(k)$, *for* $k \,\varepsilon\, \mathscr{B}$.

It is appropriate to divide the proof into two parts.

PART A: *Strong equivalence and isomorphism of algebras of bounded measurable functions.* Let ϕ be an algebraic isomorphism on $\mathfrak{M}$ to $\mathfrak{M}'$, and let $\mathscr{L}$ [or $\mathscr{L}'$] be the linear spaces of all elements of $\mathscr{B}$ [or $\mathscr{B}'$, respectively] which are finite linear combinations of characteristic functions of measurable sets. We define a mapping $\tilde{\phi}_0$ on $\mathscr{L}$ to $\mathscr{L}'$ as follows: for any mutually disjoint measurable sets E_i $(i = 1, 2, \cdots, n;\ n$ finite) in M, $\tilde{\phi}_0(\Sigma_i \alpha_i \chi_{\bar{E}_i}) = \Sigma_i \alpha_i \chi_{\phi(\bar{E}_i)}$. To see that ϕ_0 is single-valued, i. e. does not depend on the manner in which the function $f = \Sigma_i \alpha_i \chi_{\bar{E}_i}$ is represented, let $\gamma_1, \cdots, \gamma_p$ be the distinct nonzero values which f assumes on the sets (necessarily measurable) $H_0, H_1, \cdots, H_p$ of positive measure. Setting $\Lambda_1 = [i \mid \alpha_i = 0]$, $\Lambda_2 = [i \mid \alpha_i \neq 0$ and $\bar{E}_i = 0]$, $\Lambda_3 = [i \mid \alpha_i \neq 0$ and $\bar{E}_i \neq 0]$, it is clear that

$$\Sigma_i \alpha_i \chi_{\phi(\bar{E}_i)} = \Sigma_{i \,\varepsilon\, \Lambda_1} \alpha_i \chi_{\phi(\bar{E}_i)} + \Sigma_{i \,\varepsilon\, \Lambda_2} \alpha_i \chi_{\phi(\bar{E}_i)} + \Sigma_{i \,\varepsilon\, \Lambda_3} \alpha_i \chi_{\phi(\bar{E}_i)} = \Sigma_{i \,\varepsilon\, \Lambda_3} \alpha_i \chi_{\phi(\bar{E}_i)}.$$

Now putting $\Omega_j = [i \mid i \,\varepsilon\, \Lambda_3, \alpha_i = \gamma_j]$ it is plain that the Ω_j are mutually disjoint and cover Λ_3 so that

$$\Sigma_{i \,\varepsilon\, \Lambda_3} \alpha_i \chi_{\phi(E_i)} = \Sigma_j \Sigma_{i \,\varepsilon\, \Omega_j} \alpha_i \chi_{\phi(E_i)} = \Sigma_j \gamma_j \Sigma_{i \,\varepsilon\, \Omega_j} \chi_{\phi(E_i)} = \Sigma_j \gamma_j \chi_{\bigcup_{i \,\varepsilon\, \Omega_j} \phi(E_i)}$$
$$= \Sigma_j \gamma_j \chi_{\phi(\bigcup_{i \,\varepsilon\, \Omega_j} \bar{E}_i)} = \Sigma_j \gamma_j \chi_{\phi(\bar{H}_j)},$$

noting that $\bigcup_{i \,\varepsilon\, \Omega_j} \bar{E}_i = \bar{H}_j$. Since the last expression for $\tilde{\phi}_0(f)$ is independent of the manner in which f is represented, it follows that $\tilde{\phi}_0$ is single-valued.

It is easy to see that $\mathscr{L}$ and $\mathscr{L}'$ are subalgebras of $\mathscr{B}$ and $\mathscr{B}'$, and we show next that $\tilde{\phi}_0$ is a homomorphism. Let $f = \Sigma_i \alpha_i \chi_{\bar{E}_i}$ and $g = \Sigma_j \beta_j \chi_{\bar{F}_j}$ be arbitrary elements of $\mathscr{L}$. Putting $\bar{G}_{ij} = \bar{E}_i \cap \bar{F}_j$, the $\bar{G}_{ij}$ are mutually disjoint, and $\chi_{\bar{E}_i} = \chi_{\bigcup_i \bar{G}_{ij}} = \Sigma_j \chi_{\bar{G}_{ij}}$, $\chi_{\bar{F}_j} = \chi_{\bigcup_i \bar{G}_{ij}} = \Sigma_i \chi_{\bar{G}_{ij}}$. It follows that $\chi_{\phi(E_i)} = \chi_{\phi(\bigcup_j \bar{G}_{ij})} = \chi_{\bigcup_j \phi(\bar{G}_{ij})} = \Sigma_j \chi_{\phi(\bar{G}_{ij})}$ and similarly $\chi_{\phi(\bar{F}_j)} = \Sigma_i \chi_{\phi(\bar{G}_{ij})}$. Therefore $f + g = \Sigma_i \alpha_i \Sigma_j \chi_{\bar{G}_{ij}} + \Sigma_j \beta_j \Sigma_i \chi_{\bar{G}_{ij}} = \Sigma_{i,j}(\alpha_i + \beta_j)\chi_{\bar{G}_{ij}}$, and

$$\tilde{\phi}_0(f + g) = \Sigma_{i,j}(\alpha_i + \beta_j)\chi_{\phi(\bar{G}_{ij})} = \Sigma_i \alpha_i \Sigma_j \chi_{\phi(\bar{G}_{ij})} + \Sigma_j \beta_j \Sigma_i \chi_{\phi(G_{ij})}$$
$$= \Sigma_i \alpha_i \chi_{\phi(\bar{E}_i)} + \Sigma_j \beta_j \chi_{\phi(\bar{F}_j)} = \tilde{\phi}_0(f) + \tilde{\phi}_0(g).$$

It is easy to see also that $\tilde{\phi}_0(fg) = \tilde{\phi}_0(f)\tilde{\phi}(g)$ and that $\tilde{\phi}_0(\alpha f) = \alpha\tilde{\phi}_0(f)$ for any complex number α.

Now if f is as above, $\| f \|_\infty = \text{l. u. b.}_j \mid \gamma_j \mid = \| \tilde{\phi}_0(f) \|$, so that $\tilde{\phi}_0$ is an isometry. Moreover, $\tilde{\phi}_0$ is onto $\mathcal{L}'$, for if $f' = \Sigma_i \alpha_i \chi_{\bar{E}'_i}$ is an arbitrary element of $\mathcal{L}'$, putting $\bar{E}_i$ for the element of $\mathfrak{M}$ such that $\phi(\bar{E}_i) = \bar{E}_i'$ and $f = \Sigma \alpha_i \chi_{\bar{E}_i}$ it is obvious that $\tilde{\phi}_0(f) = f'$. It follows that $\tilde{\phi}_0$ can be extended to an isometry ϕ of the closure of $\mathcal{L}$ onto the closure of $\mathcal{L}'$, and it is easily seen that these closures are algebras and that ϕ is a homomorphism. To show that ϕ has the properties alleged in the theorem, it is only necessary now to show that $\mathcal{L}$ [and $\mathcal{L}'$] are dense in $\mathcal{B}$ [and $\mathcal{B}'$].

It suffices to show that the real elements of $\mathcal{L}$ are dense in the set of all real elements of $\mathcal{B}$, for if h is an arbitrary element of $\mathcal{B}$, $h = h_1 + ih_2$, where h_1 and h_2 are real, and if the h_i can be approximated by real elements of $\mathcal{L}$, then h can be approximated by complex elements; and since M is an arbitrary measure space, the density of $\mathcal{L}'$ in $\mathcal{B}'$ follows from the density of $\mathcal{L}$ in $\mathcal{B}$. Now let k be an arbitrary real element in $\mathcal{B}$, and let ϵ be an arbitrary positive number. Set $E_n = [x \mid n\epsilon < k(x) \leq (n+1)\epsilon]$ $(n = 0, \pm 1, \cdots)$. Then E_n is measurable, and is of measure zero for $|n|$ sufficiently large, say for $|n| > |m|$. Putting $k_\epsilon = \sum_{n=-m}^{m} n\epsilon\chi_{E_n}$ it is easy to conclude that $k_\epsilon \in \mathcal{L}$, and that $\| k - k_\epsilon \|_\infty \leq \epsilon$.

Remark 4.2. The mapping $\tilde{\phi}$ of Theorem 4.1 is necessarily unique. To see this, observe that any algebraic isomorphism of $\mathcal{B}$ onto $\mathcal{B}'$ preserves norms, for the norm of an element k of $\mathcal{B}$ can be defined purely algebraically, as the g. l. b. of the non-negative numbers α for which there exists an element h of $\mathcal{B}$ such that $\alpha^2 - k\bar{k} = h\bar{h}$. Thus any such isomorphism is necessarily continuous, and since it must evidently coincide on $\mathcal{L}$ with $\tilde{\phi}_0$, it is identical with $\tilde{\phi}$.

Completion of proof for Part A. Conversely now, if $\tilde{\phi}$ is an algebraic isomorphism of $\mathcal{B}$ onto $\mathcal{B}'$, and if E is any measurable set in M, then from the equation $(\chi_{\bar{E}})^2 = \chi_{\bar{E}}$ it follows that $\tilde{\phi}((\chi_{\bar{E}})^2) = \tilde{\phi}(\chi_{\bar{E}})$, which implies that $\tilde{\phi}(\chi_{\bar{E}}) = \chi_{\bar{E}'}$ for some $\bar{E}' \in \mathfrak{M}'$. We define ϕ as the mapping on $\mathfrak{M}$ to $\mathfrak{M}'$ given by the equation: $\phi(\bar{E}) = \bar{E}'$. Then ϕ is univalent, for if $\phi(\bar{E}) = \phi(\bar{F})$, $\bar{F}$ being in $\mathfrak{M}$, it results that $\chi_{\phi(\bar{E})} = \chi_{\phi(\bar{F})}$, $\tilde{\phi}(\chi_{\bar{E}}) = \tilde{\phi}(\chi_{\bar{F}})$, $\chi_{\bar{E}} = \chi_{\bar{F}}$ and so $\bar{E} = \bar{F}$. Also, ϕ is onto, for if $\bar{F}' \in \mathfrak{M}'$, then since $\tilde{\phi}$ is onto there exists $f \in \mathcal{B}$ such that $\tilde{\phi}(f) = \chi_{\bar{F}'}$; applying $\tilde{\phi}^{-1}$ to the equation $(\chi_{\bar{F}'})^2 = \chi_{\bar{F}'}$ yields the equation $f^2 = f$, from which it follows that $f = \chi_{\bar{F}}$ for some $\bar{F} \in \mathfrak{M}$, and that $\bar{F}' = \phi(\bar{F})$. Finally, ϕ is order-preserving (and hence an algebraic isomorphism) for if $\bar{F} \in \mathfrak{M}$ and $\bar{F} \leq \bar{E}$, then $\chi_{\bar{F}}\chi_{\bar{E}} = \chi_{\bar{F}}$ and applying $\tilde{\phi}$ to this equation leads to the equation $\chi_{\phi(\bar{F})}\chi_{\phi(\bar{E})} = \chi_{\phi(\bar{F})}$ which implies that $\phi(\bar{F}) \leq \phi(\bar{E})$.

PART B: *Isomorphism of algebras of bounded measurable functions and unitary equivalence of multiplication algebras.*

In this part, $\tilde{\phi}$ is a given algebraic isomorphism of $\mathfrak{B}$ onto $\mathfrak{B}'$, and it must be shown that the algebraic isomorphism Φ of $\mathfrak{A}$ onto $\mathfrak{A}'$, where Φ takes multiplication by k into multiplication by $\tilde{\phi}(k)$, is induced by an appropriate unitary transformation of $L_2(M)$ onto $L_2(M')$. We shall exhibit this unitary transformation as the product of two such transformations. The first of these is from $L_2(M)$ onto $L_2(M'')$, where M'' has the following properties: it is metrically equivalent to M, and on the other hand $M'' = (R', \mathcal{S}', s')$, where $M' = (R', \mathcal{R}', r')$, M' and M'' have the same measurable sets, and where s' and r' are absolutely continuous with respect to each other. The second transformation is then from $L_2(M'')$ onto $L_2(M')$.

Let ϕ be the isomorphism of $\mathfrak{M}$ onto $\mathfrak{M}'$ which induces $\tilde{\phi}$ as in Part A. For any measurable set E' in M' we define a measure s' by the equation $s'(E') = r(\phi^{-1}(\bar{E}'))$. It is clear that s' is countably additive on the collection of measurable sets in M', and that if $\mathcal{S}'$ is the collection of measurable sets E' such that $s'(E')$ is finite, then $M'' = (R', \mathcal{S}', s')$ is a measure space.

LEMMA 4.1.1. *A set is measurable, or measurable and σ-finite, or a null set in M', if and only if it is the same in M''.*

We observe to begin with any σ-finite measurable set H' in M' is equivalent in M' to a σ-finite measurable set in M''. For $\bar{H}'$ is then σ-finite in $\mathfrak{M}'$, so that $\bar{H}' = \phi(\bar{H})$, where $\bar{H}$ is a σ-finite element of $\mathfrak{M}$, say $\bar{H} = \bigcup_i \bar{H}_i$, where $\bar{H}_i$ has finite measure. It then follows that $\bar{H}' = \bigcup_i \bar{H}_i'$, where $\bar{H}_i'$ is a measurable set in M' such that $\bar{H}_i' = \phi(\bar{H}_i)$, or that H' is equivalent in M' to the union of the H_i'. As each H_i' is in $\mathcal{S}'$, this union is a σ-finite measurable set in M''.

It is obvious that if E' is a measurable set in M', then it is a measurable set in M''. Now if E' is a measurable set in M'', and if H' is in $\mathcal{R}'$, then H' is σ-finite in M', and hence by the preceding paragraph, $H' = \bigcup_i H_i' \cup N'$, where $H_i' \in \mathcal{S}'$ $(i = 1, 2, \cdots)$ and N' is equivalent to zero in M'. Clearly $E' \cap H' = \bigcup_i (E' \cap H_i') \cup (E' \cap N')$, and since $E' \cap H_i' \in \mathcal{S}'$ and so is measurable in M', it results that $E' \cap H'$ is measurable in M', and hence that E' is measurable in M'.

Now if G' is a σ-finite measurable set in M'', we have $G' = \bigcup_i G_i' \cup N''$, where $G_i' \in \mathcal{S}'$ and N'' is equivalent to zero in M''. Now $\bar{G}_i'$ (regarding G_i' here as a measurable set in M') is σ-finite, for $s'(\bar{G}_i') = r(\phi^{-1}(\bar{G}_i'))$ is finite, showing that $\phi^{-1}(\bar{G}_i')$ is σ-finite, and hence that $\bar{G}_i'$ is such. Hence G_i' is equivalent in M' to a σ-finite set in M', whence $G' = F' \cup N''$, where F' is σ-finite in M', and N'' is equivalent to zero in M''. It follows that to conclude the proof of the lemma, it suffices to show that a set is equivalent to zero in M' if and only if it is such in M''.

If E' is equivalent to zero in M', but not in M'', then $E' \supset W'$ where $W' \varepsilon \mathcal{S}'$ and $s'(W') \neq 0$. Now $s'(W') = r'(\phi^{-1}(\bar{W}'))$ (here $\bar{W}' = \theta_{M'}(W')$, i. e. $\bar{W}'$ is the residue class containing W' relative to M') which shows that $\phi^{-1}(\bar{W}') \neq 0$ and hence $\bar{W}' \neq 0$. This contradicts the relations $\bar{W} \leq \bar{E}'$ and $\bar{E}' = 0$. If on the other hand E'' is equivalent to zero in M'', but not in M', then $E'' \supset W'$, where $W' \varepsilon \mathcal{R}'$ and $\bar{W}' \neq 0$. Now E'' is measurable in M'' and $\bar{E}'' \geq \bar{W}'$ ($\bar{E}'' = \theta_{M'}(E'')$), so $s'(\bar{E}'') \geq s'(\bar{W}')$. It follows that $0 \geq r(\phi^{-1}(\bar{W}'))$, whence $\phi^{-1}(\bar{W}') = 0$, and $\bar{W}' = 0$, a contradiction.

Resumption of proof of Part B *of theorem.* Our next step is to set up a unitary transformation from $B_2(M)$ onto $B_2(M'')$. This transformation will be seen to be the unique extension to $B_2(M)$ of the contraction of $\tilde{\phi}$ to $\mathfrak{B} \cap B_2(M)$.

Let f be an arbitrary element of $L_2(M) \cap L_\infty(M)$, and let $f'' = \phi(f)$. Since f'' is measurable with respect to M', and since every set measurable in M' is likewise measurable in M'', f'' is measurable with respect to M''. To evaluate the norm of f'' in $L_2(M'')$, we note that

$$\int_{R'} |f''(x)|^2 \, ds'(x) = -\int_0^\infty \lambda^2 \, ds'(S_\lambda'),$$

where $S_\lambda' = [x \mid |f''(x')| > \lambda]$. Now $\tilde{\phi}$ preserves order, on the subspace of real functions in $\mathfrak{B}$ (where $p \geq q$ means that p-q is n. e. non-negative), for $p \geq q$, where p and q are real functions in $\mathfrak{B}$, holds if and only if $p = q + r^2$, where r is a real function in $\mathfrak{B}$; and this implies

$$\tilde{\phi}(p) = \tilde{\phi}(q) + (\tilde{\phi}(r))^2, \ \phi(p), \ \phi(q), \text{ and } \phi(r)$$

being real because ϕ preserves complex conjugation. Putting

$$S_\lambda = [x \mid |f(x)| > \lambda],$$

it is clear that $f\bar{f} \geq \lambda^2 \chi_{S_\lambda}$ and that $f''\bar{f}'' \geq \lambda^2 \chi_{S_\lambda'}$. Applying $\tilde{\phi}$ to the first of these inequalities it results that $f''\bar{f}'' \geq \lambda^2 \chi_{\phi(S_\lambda)}$ which implies that $\phi(\bar{S}_\lambda) \leq \bar{S}_\lambda'$. Applying $\tilde{\phi}^{-1}$ to the second inequality yields the inequality $f\bar{f} \geq \lambda^2 \chi_{\phi^{-1}(S_\lambda')}$ which shows that $\phi^{-1}(\bar{S}_\lambda') \leq \bar{S}_\lambda$, and hence that $\bar{S}_\lambda' \leq \phi(\bar{S}_\lambda)$. It follows that $\phi(\bar{S}_\lambda) = \bar{S}_\lambda'$. Now $s'(\phi(\bar{E})) = r(\bar{E})$ for any $\bar{E} \varepsilon \mathfrak{M}$, and hence

$$\int_{R'} |f''(x)|^2 \, ds'(x) = -\int_0^\infty \lambda^2 \, dr(S_\lambda) = \int_R |f(x)|^2 \, dr(x).$$

Thus $\tilde{\phi}$ maps $\mathfrak{B} \cap B_2(M)$ into $\mathfrak{B}' \cap B_2(M')$ in a manner which preserves the B_2 norm. Moreover, if f'' is any element of $\mathfrak{B}' \cap B_2(M')$, $\tilde{\phi}^{-1}(f'')$ is in $\mathfrak{B}$, and by the last line of the preceding paragraph (which is valid for arbitrary $f \varepsilon \mathfrak{B}$), it is also in $B_2(M)$. Thus $\tilde{\phi}$ maps $\mathfrak{B} \cap B_2(M)$ onto $\mathfrak{B}' \cap B_2(M')$. Since $\mathfrak{B} \cap B_2(M)$ and $\mathfrak{B}' \cap B_2(M')$ are dense in $B_2(M)$ and $B_2(M')$ respectively, we can conclude that there is a unique isometric linear map V of $B_2(M)$ onto $B_2(M')$ (i. e. V is unitary) such that $Vf = \tilde{\phi}(f)$, for $f \varepsilon \mathfrak{B} \cap B_2(M)$.

Next we construct the unitary map mentioned above, of $B_2(M'')$ onto $B_2(M')$. If E' is any σ-finite measurable set in M', it is easy to deduce from the Radon-Nikodym theorem (applied to the spaces whose set is E' and whose measurable sets are the M'-measurable sets of E', to which the same measures as those in M' and M'' are given) that there exists a measurable function $h_{E'}$ on M' which is n. e. positive on E', vanishes on the complement of E', and such that for any measurable subset F' of E',

$$s'(F') = \int_{F'} h_{E'}(x')\,dr'(x')\,;\; r'(F') = \int_{F'} \{h_{E'}(x')\}^{-1}\,ds'(x').$$

We observe that if H' is any σ-finite measurable set in M' which contains E', then $h_{H'}(x') = h_{E'}(x')$ n. e. on E', for

$$\int_{F'} h_{H'}(x')\,ds'(x') = \int_{F'} h_{E'}(x')\,ds'(x'),$$

for any measurable subset F' of E'.

We now define a unitary map W on $L_2(M'')$ to $L_2(M')$ as follows: if $f \varepsilon L_2(M'')$, if E' is a σ-finite measurable set outside of which f vanishes, then $(Wf)(x') = f(x')\{h_{E'}(x')\}^{-\frac{1}{2}}$ for $x \varepsilon E'$ and $(Wf)(x') = 0$ if $x' \notin E'$. It is easy to verify that Wf is measurable on M', and (up to equivalence) is independent of the choice, within the stated restrictions, of E'; and $Wf \varepsilon L_2(M')$, in fact

$$\| Wf \|_2{}^2 = \int_{E'} | f(x') |^2 \{h_{E'}(x)\}^{-1} dr'(x') = \int_{E'} | f(x') |^2 ds'(x) = \| f \|_2^2$$

so that W is an isometry. To see that W is linear, let g be arbitrary in $L_2(M'')$ and let α and β be arbitrary constants. Then if F' is a σ-finite measurable set outside of which g vanishes, and if $G' = E' \cup F'$, we have

$$\begin{aligned}\{W(\alpha f + \beta g)\}(x') &= \{\alpha f(x') + \beta g(x')\}\{h_{G'}(x')\}^{-\frac{1}{2}} \\ &= \alpha f(x')\{h_{G'}(x')\}^{-\frac{1}{2}} + \beta g(x')\{h_{G'}(x')\}^{-\frac{1}{2}} \\ &= \alpha(Wf)(x') + \beta(Wg)(x').\end{aligned}$$

Thus W is unitary on $L_2(M'')$ to $L_2(M')$ and the product $U = WV$ is unitary on $L_2(M)$ to $L_2(M')$. It remains only to show that the algebraic isomorphism of $\mathcal{A}$ induced by U is actually Φ. That is, it must be shown that if $k \varepsilon \mathcal{B}$, then $UT_k = T_{\tilde{\phi}(k)}U$, where T_m denotes the operation of multiplication by the function m on the space of square-integrable functions over the measure space on which m is defined. Thus, the equation $UT_kf = T_{\tilde{\phi}(k)}Uf$ must be checked, for all $f \varepsilon L_2(M)$; but since U, T_k, and $T_{\tilde{\phi}(k)}$ are bounded operators, it is sufficient to check the equation for a dense set of f's, e. g. for $f \varepsilon L_2(M) \cap L_\infty(M)$. Now for such an f, $(UT_kf)(x') = (Up)(x')$, where $p(x) = k(x)f(x)$; $= (WVp)(x') = (Wq)(x')$, where $q = \tilde{\phi}(p) = \tilde{\phi}(k)\tilde{\phi}(f)$; $= q(x')\{h_{E'}(x')\}^{-\frac{1}{2}}$ on E' and 0 elsewhere, where E' is any σ-finite measurable set outside of which q vanishes; $= (\tilde{\phi}(k))(x')(\tilde{\phi}(f))(x')\{h_{E'}(x')\}^{-\frac{1}{2}}$ on E' and 0 elsewhere. On the other hand, $(T_{\tilde{\phi}(k)}Uf)(x') = (T_{\tilde{\phi}(k)}WVf)(x')$ $= (T_{\tilde{\phi}(k)}W\tilde{\phi}(f))(x') = (T_{\tilde{\phi}(k)}t)(x')$, where $t(x') = (\tilde{\phi}(f))(x')\{h_F'(x')\}^{-\frac{1}{2}}$ on F' and 0 elsewhere, F' being any σ-finite measurable set outside of which $\tilde{\phi}(f')$ vanishes; $= \tilde{\phi}(k)(x')t(x') = (\tilde{\phi}(k))(x')(\tilde{\phi}(f))(x')\{h_{F'}(x')\}^{-\frac{1}{2}}$ on F' and 0 elsewhere. Since $h_{E'}$ and $h_{F'}$ coincide n. e. on $E' \cap F'$, outside of which q vanishes, it results that $(UT_kf)(x') = T_{\tilde{\phi}(k)}Uf)(x')$ n. e. and the proof is complete.

The next theorem relates weak equivalence of measure spaces to varieties of isomorphism of function-spaces over measure spaces. Such equivalence could also be related to unitary equivalence of the algebras of multiplications by bounded measurable functions vanishing outside of σ-finite sets, in a fashion similar to that described in the statement and proof of the preceding theorem.

THEOREM 4. 2. *If the spaces X and X' of real-valued α-th-power integrable functions over two measure spaces M and M' are isomorphic, as partially ordered linear spaces ($\alpha \geqq 1$), then M and M' are weakly equivalent.*

If on the other hand M and M' are weakly equivalent, then there is an isometric isomorphism between X and X', as partially ordered Banach spaces.

We show first that M and M' are weakly equivalent if there is an isomorphism between X and X' as partially ordered linear spaces. Let ψ be such an isomorphism, i. e. an order-preserving linear map of X onto X'. For any numerical function f on a measure space, let $S(f)$ denote the set of points in the space where the function is nonzero. If $\bar{E}$ is any element of the σ-finite measure ring $\mathcal{J}$ of M, we define $\phi(\bar{E})$ as follows: if f is a non-negative element of X such that $\bar{S}(f)$ $(= \theta(S(f))) = \bar{E}$, then $\phi(\bar{E}) = \bar{S}(\psi(f))$. It is easy to see that such a function f always exists, so that $\phi(\bar{E})$ exists for all $\bar{E} \varepsilon \mathcal{J}$, and we show next that ϕ is single-valued.

Suppose that f and g are non-negative elements of X such that $\bar{S}(f) = \bar{S}(g) = \bar{E}$. Then setting $k = f \cup g$ and $h = f \cap g$ (where $\cup$ and $\cap$ denote the l. u. b. and g. l. b. respectively in X) it results that k and h are in X and that $\bar{S}(k) = \bar{S}(h) = \bar{E}$. Plainly, $\psi(k) \geq \psi(h)$, so that $\bar{S}(\psi(k)) \geq \bar{S}(\psi(h))$. On the other hand, if $\bar{S}(\psi(k)) \neq \bar{S}(\psi(h))$, then setting p' for a nonzero element of X' such that $\psi(k) \geq p'$ and $p' \cap \psi(h) = 0$ (e. g. $p'(x') = (\psi(k))(x')$ for $x' \varepsilon S(\psi(k)) - S(\psi(h))$ and $= 0$ otherwise), and $p = \psi^{-1}(p')$, it follows that $p \leq k$ and $p \cap h = 0$. These last relations imply the contradiction $p = 0$, and hence $\bar{S}(\psi(k)) = \bar{S}(\psi(h))$. Now from the inequalities $k \geq f \geq h$ it follows that $\bar{S}(\psi(k)) \geq \bar{S}(\psi(f)) \geq \bar{S}(\psi(h))$, and hence we can conclude that $\bar{S}(\psi(f)) = \bar{S}(\psi(k))$. Since k is a symmetric function of f and g, this equation implies that $\bar{S}(\psi(f)) = \bar{S}(\psi(g))$, i. e. ϕ is single-valued.

To complete the proof that M and M' are weakly equivalent it suffices to show that ϕ is onto, univalent, and order-preserving. If $\bar{E}'$ is in the σ-finite measure ring $\mathfrak{F}'$ of M', then there exists an element f' in X' such that $\bar{S}(f') = \bar{E}'$, and hence $\bar{E}' = \phi(\bar{E})$, where $\bar{E} = \bar{S}(\psi^{-1}(f'))$; i. e. ϕ is onto. To see that ϕ is univalent, let $\bar{F}$ and $\bar{G}$ be elements of $\mathfrak{F}$ such that $\phi(\bar{F}) = \phi(\bar{G})$. Putting f' for an element of X' for which $\bar{S}(f') = \phi(\bar{F}) = \phi(\bar{G})$, it follows from the definition of ϕ that if $f = \psi^{-1}(f')$, then $\bar{S}(f) = \bar{F} = \bar{G}$. Finally, if $\bar{F}$ and $\bar{G}$ are elements of $\mathfrak{F}$ such that $\bar{F} \leq \bar{G}$, then it is easily seen that there exist non-negative elements f and g of X for which $\bar{S}(f) = \bar{F}$, $\bar{S}(g) = \bar{G}$, and $f \leq g$. The last inequality implies that $\psi(f) \leq \psi(g)$, from which it results that $\bar{S}(\psi(f)) \leq \bar{S}(\psi(g))$, i. e. $\phi(\bar{F}) \leq \phi(\bar{G})$.

Now let ϕ be a weak equivalence between M and M', from $\mathfrak{F}$ onto $\mathfrak{F}'$. By Theorem 3. 4, M is metrically equivalent to a measure space P which is a direct sum of finite measure spaces, and hence localizable. It is easy to see from the proof of Theorem 3. 4 that there can be assumed to exist an isometric order-preserving linear map A of X onto $RL_\alpha(P)$. Since M is an arbitrary measure space, there is a map B of the same type as A on X' to $RL_\alpha(P')$, where P' is a localizable measure space. Since M and M' are weakly equivalent, so likewise are P and P', and hence, by Theorem 3. 3, P and P' are strongly equivalent.

It is not difficult to verify that the unitary transformation U of the proof of Theorem 4. 1 is order-preserving. Hence there exists an isometric linear order-preserving map U of $RL_2(P)$ onto $RL_2(P')$. It is easy to conclude that $B^{-1}UA$ is an isometric order-preserving linear map of X onto X', and so the proof of the theorem is complete in the case $\alpha = 2$. For general α the theorem can be concluded in a similar fashion as follows. Putting ξ for the algebraic isomorphism of $B_\infty(P)$ onto $B_\infty(P')$ induced by the strong equivalence between P and P', and noting that in the present case $s' = r'$ (in the notation of the proof of Theorem 4. 1), it is straightforward to verify by the procedure of that proof, that the contraction of ξ to $B_\infty(P) \cap B_a(P)$ can be extended to an isometric, linear, order-preserving map U of $B_a(P)$ onto $B_a(P')$. Then, as above, $B^{-1}UA$ is the required map of X onto X'.

We conclude this section by characterizing metric equivalence in terms of the induced correspondence between certain function spaces.

THEOREM 4. 3. *Two measure spaces M and M' are metrically equivalent if and only if there is an algebraic isomorphism between the algebras of bounded real-valued measurable functions vanishing at infinity on M and M' which preserves the integrals of idempotents in the algebras.*

Let $\mathcal{A}$ and $\mathcal{A}'$ be the algebras of bounded real-valued measurable functions vanishing at infinity on $M = (R, \mathcal{R}, r)$ and $M' = (R', \mathcal{R}', r')$, and let ϕ be an algebraic isomorphism from $\mathcal{A}$ onto $\mathcal{A}'$ which preserves the integrals of idempotents in $\mathcal{A}$. It is easy to see that an idempotent f in $\mathcal{A}$ is the characteristic function of a chunk in M, and that conversely every such function is an idempotent in $\mathcal{A}$. It follows that if E is any chunk in M, there is a chunk E' in M' such that $\phi(\chi_{\bar{E}}) = \chi_{\bar{E}'}$, and $r(E) = r'(E')$. Now let $\mathcal{D}$ and $\mathcal{D}'$ be the subringc of $\mathcal{F}$ and $\mathcal{F}'$ respectively which consist of the elements of finite measure, and put $\tilde{\phi}$ for the mapping on $\mathcal{D}$ to $\mathcal{D}'$ determined by the equation $\tilde{\phi}(\bar{E}) = \bar{E}'$, where $\phi(\chi_{\bar{E}}) = \chi_{\bar{E}'}$, for $\bar{E} \varepsilon \mathcal{D}$. It is easy to see that ϕ is a measure-preserving ring isomorphism of $\mathcal{D}$ onto $\mathcal{D}'$.

Now let M and M' be metrically equivalent, and let η be a measure-preserving ring isomorphism of the finite measure ring $\mathcal{D}$ of M onto the corresponding ring $\mathcal{D}'$ of M'. Put $\mathcal{L}$ and $\mathcal{L}'$ for the subalgebras of $\mathcal{A}$ and $\mathcal{A}'$ consisting of all real finite linear combinations of characteristic functions of chunks in M and M' respectively. It is easy to see by the procedure used in the proof of Part A of Theorem 4. 1, that the mapping $\phi_0 : \Sigma_i \alpha_i \chi_{\bar{E}_i} \rightarrow \Sigma_i \alpha_i \chi_{\eta(\bar{E}_i)}$, where $\bar{E}_i \varepsilon \mathcal{D}$, i ranges over a finite set, and the α_i are real, is a (single-valued) algebraic isomorphism of $\mathcal{L}$ onto $\mathcal{L}'$. Moreover, $\| \phi_0(f) \|_\infty \leq \| f \|_\infty$, for $f \varepsilon \mathcal{L}$. It is not difficult to verify that $\mathcal{L}$ and $\mathcal{L}'$ are dense in $\mathcal{A}$ and $\mathcal{A}'$ respectively (with respect to the B_∞ norm topology), and it follows that ϕ_0 can be extended to an algebraic isomorphism of $\mathcal{A}$ onto $\mathcal{A}'$. It is obvious that ϕ preserves the integrals of idempotents in $\mathcal{A}$.

We need the following definition before stating our final characterization of metric equivalence.

DEFINITION 4. 2. Let M and M' be measure spaces. A mapping ϕ on $L_2(M)$ to $L_2(M')$ is called *multiplicative* if whenever elements f, g, and h

of $L_2(M)$ are such that $f(x)g(x) = h(x)$ n. e. on M, then the equation $f'(x')g'(x') = h'(x')$ holds n. e. on M', where f', g', and h' are the transforms of f, g, and h respectively, under ϕ.

THEOREM 4. 4. *Two measure spaces M and M' are metrically equivalent if and only if there is a multiplicative unitary transformation from $L_2(M)$ onto $L_2(M')$.*

We treat first the "only if" part of the theorem. Let η be a measure-preserving ring isomorphism of the finite measure ring $\mathscr{D}$ of $M = (R, \mathscr{R}, r)$ onto the corresponding ring $\mathscr{D}'$ of M', and let $\mathscr{L}$ and $\mathscr{L}'$ be the algebras of all complex finite linear combinations of characteristic functions of chunks in M and M' respectively. Then as in the proof of Theorem 4. 1, the mapping: $\phi_0 : \Sigma_i \alpha_i \chi_{\bar{E}_i} \rightarrow \Sigma_i \alpha_i \chi_{\eta(\bar{E}_i)}$, where $\bar{E}_i \varepsilon \mathscr{D}$, i ranges over a finite set, and the α_i are complex numbers, is a (single-valued) algebraic isomorphism of $\mathscr{L}$ onto $\mathscr{L}'$. Just as in the proof of the preceding theorem ϕ_0 can be extended to an algebraic isomorphism ϕ of the algebra $\mathscr{B}$ of all complex-valued measurable functions vanishing at infinity on M, onto the corresponding algebra $\mathscr{B}'$ on M' (both algebras modulo the ideals of functions equivalent to zero). Plainly $\mathscr{B} \cap B_2(M)$ and $\mathscr{B}' \cap B_2(M')$ are dense subsets of $B_2(M)$ and $B_2(M')$ respectively, and it can be verified as in the proof of Theorem 4. 1, that ϕ is isometric, relative to the B_2 norm, from $\mathscr{B} \cap B_2(M)$ to $\mathscr{B}' \cap B_2(M')$. It follows that the contraction of ϕ to $\mathscr{B} \cap B_2(M)$ can be uniquely extended to a unitary transformation U from $B_2(M)$ onto $B_2(M')$.

To complete the proof of this part of the theorem, it suffices to show that U is multiplicative. Let f, g, and h be elements of $L_2(M)$ such that $f(x)g(x) = h(x)$ n. e. on M, and set

$$F_n = [x \mid | f(x) | \leq n], \quad G_n = [x \mid | g(x) | \leq n], \quad H_n = [x \mid | h(x) \leq n],$$
$$\text{and } K_n = F_n \cap G_n \cap H_n \qquad (n = 1, 2, \cdots).$$

Plainly each of $\bigcup_n F_n$, $\bigcup_n G_n$, $\bigcup_n H_n$ is equivalent to R, and hence so also is $\bigcup_n K_n$. Now let $f_n = f\chi_{K_n}$, $g_n = g\chi_{K_n}$, and $h_n = h\chi_{K_n}$. Clearly, $f_n \rightarrow f$, $g_n \rightarrow g$, and $h_n \rightarrow h$, in $L_2(M)$, as $n \rightarrow \infty$; $f_n(x)g_n(x) = h_n(x)$ n. e. on M; and f_n, g_n, and h_n are in $\mathscr{B}$. It results in particular that $Uf_n = \phi(f_n) = f_n'$, $Ug_n = \phi(g_n) = g_n'$, and $Uh_n = \phi(h_n) = h_n'$. Now let $\{n_i\}$ be a subsequence of the positive integers such that $f_{n_i}'(x') \rightarrow f'(x')$, $g_{n_i}'(x') \rightarrow g'(x')$, and $h_{n_i}'(x') \rightarrow h'(x')$ n. e. on M', where $f' = Uf$, $g' = Ug$, and $h' = Uh$. Observing that $h_n' = \phi(f_n g_n) = \phi(f_n)\phi(g_n) = f_n' g_n'$, we have n. e. on

$$M' h'(x') = \lim_i h_{n_i}'(x') = \lim_i f_{n_i}'(x') g_{n_i}'(x')$$
$$= \lim_i f_{n_i}'(x') \lim_i g_{n_i}'(x') = f'(x')g'(x').$$

Now conversely let T be a multiplicative unitary transformation from $L_2(M)$ onto $L_2(M')$. If E is a chunk in M, then putting k for the characteristic function of E, $T(k^2) = T(k) = (T(k))^2$, so that $T(k)$ is a characteristic function of a measurable set E', and as T is unitary, $\| k \|_2 = \| k' \|_2$, where k' is the characteristic function of E', whence $r(E) = r'(E')$. Putting $\bar{E}' = \psi(\bar{E})$, it follows readily by a device previously employed that ψ is order-preserving, and ψ is clearly one-to-one and onto the finite measure ring of M'. It follows that ψ is an algebraic isomorphism of the finite measure ring of M onto that of M', and hence is a metric equivalence of M with M'.

5. Localizable spaces.

We show in the present section that localizable spaces are characterized by any of a number of important measure-theoretic properties. These results, together with those of the preceding sections, suggests that, at least as regards applications to many aspects of functional analysis, the localizable spaces constitute the most significant general class of measure spaces properly containing the finite ones, and are of the maximum generality consistent with usefulness in any large sector of abstract analysis.

Before stating the main theorem of this section we make the following

DEFINITIONS 5.1. A collection $\mathcal{K}$ of chunks in a measure space [or of elements of the finite measure ring $\mathfrak{D}$ of the space] is called a *conditional σ-ideal* if it is an ideal in the ring of measurable sets [or in $\mathfrak{M}$], and if any chunk [or element of $\mathfrak{D}$] which is a countable union of elements of $\mathcal{K}$ is itself in $\mathcal{K}$. An ideal $\mathcal{L}$ of chunks [or of elements of $\mathfrak{D}$] is said to be *localized* in L [or $\bar{L}$] if there is a measurable set L [or element $\bar{L}$ of the measure ring] such that every element of $\mathcal{L}$ is equivalent to a subset of L, and conversely every chunk contained in L is equivalent to an element of $\mathcal{L}$ [or such that $\mathcal{L} = [\bar{L} \cap \bar{D} \mid \bar{D} \in \mathfrak{D}]$].

DEFINITIONS 5.2. The *Radon-Nikodym theorem is* said to be *vaild for a space* $M = (R, \mathfrak{R}, r)$ if whenever $M' = (R, \mathfrak{R}, r')$ is a measure space such that $r'(E) \leq \alpha r(E)$ for $E \in \mathfrak{R}$, α being a constant, then there exists a measurable function f on M such that for $E \in \mathfrak{R}$, $r'(E) = \int_E f(x) dr(x)$. The *Riesz representation theorem* for linear functionals on L_1 *is* said to be *valid* for M if for every continuous linear functional ϕ on $L_1(M)$ there is a (fixed) bounded measurable function k on M such that

$$\phi(f) = \int_R k(x) f(x) dr(x) \, (f \in L_1(M)).$$

Remarks 5.1. It is easy to see that if a collection $\mathcal{K}$ of chunks is a conditional σ-ideal, then the same is true of $\theta(\mathcal{K})$; and that $\mathcal{K}$ is localized in k if and only if $\theta(\mathcal{K})$ is localized in $\theta(K)$.

The form of the Radon-Nikodym theorem given above differs somewhat from the usual version. It is easy to see however by a well-known device that the result for the case when r' is merely absolutely continuous relative to r follows from that for the special case described.

The following theorem further justifies our term "localizable," by showing that such spaces are precisely those in which every conditional σ-ideal of chunks is localized.

THEOREM 5. 1. *Any of the following conditions on a measure space* $M = (R, \mathcal{R}, r)$ *implies all of the other conditions.*

a. M is localizable.

b. M is strongly equivalent to a direct sum of finite measure spaces.

c. The Radon-Nikodym theorem is valid for M.

d. The multiplication algebra of M is maximal abelian in the algebra of all bounded linear operators on $L_2(M)$.

e. Every conditional σ-ideal of chunks is localized.

f. The Riesz representation theorem for linear functionals on L_1 is valid for M.

That a and b are equivalent was proved in Section **3** and is stated here for convenience of reference. In order to prove the theorem, it therefore suffices to show that a and e are equivalent and to establish the following cycle of implications: $e \Rightarrow d$, $d \Rightarrow f$, $f \Rightarrow c$, and $c \Rightarrow e$.

A: $a \Rightarrow e$. Let $\mathcal{K}$ be a conditional σ-ideal in the finite measure ring $\mathcal{D}$ of the localizable space $M = (R, \mathcal{R}, r)$. Setting $\bar{K} = \text{l. u. b.}_{\bar{E} \varepsilon \mathcal{K}} \bar{E}$, then every element of $\mathcal{K}$ is bounded by $\bar{K}$, and it remains to show that if $\bar{F} \leq \bar{K}$ and $\bar{F} \varepsilon \mathcal{D}$, then $\bar{F} \varepsilon \mathcal{K}$. Now for such an $\bar{F}$ we have

$$\bar{F} = \bar{F} \cap \bar{K} = \bar{F} \cap \text{l. u. b.}_{\bar{E} \varepsilon \mathcal{K}} \bar{E} = \text{l. u. b.}_{\bar{E} \varepsilon \mathcal{K}} (\bar{E} \cap \bar{F}).$$

Choosing $\bar{E}_i$ in $\mathcal{K}$ such that $r(\bar{E}_i \cap \bar{F}) \rightarrow \text{l. u. b.}_{\bar{E} \varepsilon \mathcal{K}} r(\bar{E} \cap \bar{F})$, it is easy to verify that $\bar{F} \leq \bigcup_i \bar{E}_i$, and hence $\bar{F} \varepsilon \mathcal{K}$.

B: $e \Rightarrow a$. Let $\mathcal{L}$ be any subset of the measure ring $\mathfrak{M}$ of the measure space $M = (R, \mathcal{R}, r)$, for which e holds; it must be shown that $\mathcal{L}$ has a l. u. b. in $\mathfrak{M}$. Set $\mathcal{L}_0$ for the set of all elements of $\mathcal{D}$, the finite measure ring of $\mathfrak{M}$, which are bounded by some element of $\mathcal{L}$, and let $\mathcal{K}$ be the conditional σ-ideal generated by $\mathcal{L}_0$. On the one hand $\mathcal{K}$ is the intersection of all conditional σ-ideals in $\mathcal{D}$ which contain $\mathcal{L}_0$ and so exists; and on the other hand, it is easy to verify that $\mathcal{K}$ consists of all elements of $\mathcal{D}$ which are countable unions of elements of $\mathcal{D}$ which are bounded by elements of $\mathcal{L}$. Now let $\mathcal{K}$ be localized in $\bar{K}$. Then $\bar{K}$ is an upper bound for the elements of $\mathcal{L}$, since it is an upper bound for the elements of $\mathcal{D}$ which are bounded by the elements of $\mathcal{L}$, and since by Lemma 3. 3. 1 any element of $\mathcal{L}$ is the l. u. b. of the elements of $\mathcal{D}$ which it bounds. Furthermore, $\bar{K}$ is the l. u. b., for if $\bar{K}'$ is any other upper bound, it is clear from the form of $\mathcal{K}$ that $\bar{K}'$ is an upper bound for $\mathcal{K}$. This implies that $\bar{K} - \bar{K}'$ is disjoint from every element of $\mathcal{K}$, but $(\bar{K} - \bar{K}') \cap \bar{F}$ is in $\mathcal{K}$ for every $\bar{F} \in \mathcal{D}$, being bounded by $\bar{K}$, and hence $(\bar{K} - \bar{K}') \cap \bar{F} = 0$ for all such $\bar{F}$, whence $\bar{K} - \bar{K}' = 0$, i. e. $\bar{K}' \geq \bar{K}$.

C: $e \Rightarrow d$. If M is a measure space for which e holds, then as just proved, M is localizable, and hence by Theorem 3. 4 strongly equivalent to a direct sum M' of finite measure spaces. By Theorem 4. 1 the multiplication algebra of M is unitarily equivalent to the multiplication algebra of M', and hence it suffices to show that d holds for any direct sum of finite measure spaces, say for $M = \sum_{\xi \in \Xi} \oplus M_\xi$, where $M_\xi = (R_\xi, \mathfrak{R}_\xi, r_\xi)$ is finite.

Let U be a bounded linear operator on $L_2(M)$ which commutes with all multiplications T_k by bounded measurable functions k on M. Let χ_ξ be the characteristic function of R_ξ and let P_ξ be the operation of multiplication by χ_ξ. Then $T_k U \chi_\xi = U T_k \chi_\xi$, for $\xi \in \Xi$ and $k \in L_\infty(M)$, and putting $k = \chi_\xi$, it follows that $P_\xi U \chi_\xi = U \chi_\xi$, i. e. $U \chi_\xi$ vanishes n. e. outside of R_ξ. Now if g is any element of $L_2(M) \cap L_\infty(M)$ which vanishes outside of R_ξ, then $Ug = U T_g \chi_\xi = T_g U \chi_\xi = U_\xi g$, where U_ξ is the operation of multiplication by $U \chi_\xi$ on the domain $L_2(M) \cap L_\infty(M)$. Clearly $\| U_\xi g \| = \| Ug \| \leq \| U \| \| g \|$ for such g, and in particular, putting g for the characteristic function of the subset Z of R_ξ on which $U \chi_\xi$ exceeds $\| U \| + 1$ in absolute value, we have $(\| U \| + 1)(r(Z))^{\frac{1}{2}} \leq \| U_\xi g \| \leq \| U \| (r(Z))^{\frac{1}{2}}$. It follows that $r(Z) = 0$, which implies that $U \chi_\xi$ is bounded n. e. by $\| U \| + 1$ on R_ξ.

Now let h be the function on R which coincides on R_ξ with $U \chi_\xi$. It is plain that h is measurable, essentially bounded by $\| U \| + 1$, and that if g is in $L_2(M)$ and vanishes outside of some R_ξ, then $Ug = T_h g$. The last equation holds also if g vanishes outside of the union of a finite number of the R_ξ, and thus holds for a set of g's which is dense in $L_2(M)$. Since U and T_h are bounded operators, it results that they are the same.

D: $d \Rightarrow f$. The following proof is related to the proofs of the Radon-Nikodym theorem which have been given by von Neumann [8] and Stone [6]. Let d hold for the measure space $M = (R, \mathfrak{R}, r)$ and let ϕ be a continuous linear functional on $L_1(M)$ of bound α. We define a hermitian bilinear form Φ on $L_2(M)$ as follows: if $f, g \varepsilon L_2(M)$, $\Phi(f, g) = \phi(f\bar{g})$. Then $|\Phi(f, g)| \leq \alpha \| f\bar{g} \|_1 \leq \alpha \| f \|_2 \| g \|_2$, so that for each fixed g, $\Phi(f, g)$ is a bounded linear functional of f, of bound not greater than $\alpha \| g \|_2$. By the representation theorem for bounded linear functionals on Hilbert space, there exists an element $H(g)$ in $L_2(M)$ such that $\Phi(f, g) = (f, H(g))$, and $\| H(g) \|_2 \leq \alpha \| g \|_2$. It follows easily that H is a bounded linear operator on $L_2(M)$.

Now H^* commutes with all operators on $L_2(M)$ of the form T_k, where T_k is multiplication by the bounded measurable function k, for if f and g are arbitrary in $L_2(M)$, then $(H^* T_k f, g) = (T_k f, Hg) = \Phi(T_k f, g) = \phi(kf\bar{g})$ $= \Phi(f, T_{\bar{k}} g) = (f, HT_{\bar{k}} g)$, and noting that $(T_k)^* = T_{\bar{k}}$, the last expression is $(T_k H^* f, g)$. Since d holds for M, $H = T_p$ for some p in $L_\infty(M)$. It results now from the fact that the product of two arbitrary functions in $L_2(M)$ is an arbitrary function in $L_1(M)$ that for all $f \varepsilon L_1(M)$, $\phi(f) = \int_R p(x) f(x)\, dr(x)$.

E: $f \Rightarrow c$. Now let f hold in the measure space $M = (R, \mathfrak{R}, r)$ and let $M' = (R, \mathfrak{R}, r')$ be a measure space for which $r'(E) \leqq \alpha r(E)$, $E \varepsilon \mathfrak{R}$, α being a fixed positive number. We define a functional ϕ on $L_1(M)$ by the equatioin $\phi(f) = \int_R f(x)\, dr'(x)$, $f \varepsilon L_1(M)$; the indicated integral exists, and in fact

$$\int_R | f(x) |\, dr'(x) = \int_0^\infty \lambda dr'[x \mid | f(x) | \leq \lambda]$$

$$\leq \alpha \int_0^\infty \lambda dr[x \mid | f(x) | \leq \lambda] = \alpha \int_R | f(x) |\, dr(x),$$

i. e. $|\phi(f)| \leq \alpha \| f \|_1$. As f holds, there exists $p \varepsilon L_\infty(M)$ such that $\phi(f) = \int_R p(x) f(x)\, dr(x)$ for $f \varepsilon L_1(M)$, and it is readily shown that $0 \leq p(x) \leq \alpha$ n. e. on M. Thus for $f \varepsilon L_1(M)$,

$$\int_R f(x)\, dr'(x) = \int_R p(x) f(x)\, dr(x),$$

and putting f equal to the characteristic function of a set in $\mathfrak{R}$, c follows.

F: $c \Rightarrow e$. Let $M = (R, \mathfrak{R}, r)$ satisfy c, and let $\mathcal{K}$ be a conditional σ-ideal of chunks in M. We show that $\mathcal{K}$ is localized essentially by constructing a kind of quotient space of M with respect to $\mathcal{K}$, and deriving from this quotient a measure which is absolutely continuous with respect to r.

If $E \varepsilon \mathcal{R}$, the least upper bound of $\bar{E} \cap \bar{F}$ as $\bar{F}$ varies over $\theta(\mathcal{K})$ exists and is in $\theta(\mathcal{K})$, for if $\bar{F}_i$ $(i = 1, 2, \cdots)$ is a sequence of elements of $\mathcal{K}$ for which $\lim_i r(\bar{E} \cap \bar{F}_i) = \text{l. u. b.}_{F \varepsilon \theta(\mathcal{K})} r(\bar{E} \cap \bar{F})$, then it is easily seen that $\bar{E} \cap \bigcup_i \bar{F}_i$ is the least upper bound in question, and it is clear from the circumstance that $\theta(\mathcal{K})$ is a conditional σ-ideal that it is in $\theta(\mathcal{K})$. We now put $\psi(\bar{E})$ for that least upper bound and define a function r' on R as follows: $r'(E) = r(\psi(\bar{E}))$, $E \varepsilon \mathcal{R}$. Evidently $r'(E) \leq r(E)$, for $E \varepsilon \mathcal{R}$, and we show next, employing a simplification of our original argument due to the referee, that r' is a countably additive measure on $\mathcal{R}$.

Let $\{E_i\}$ be a sequence of mutually disjoint elements of $\mathcal{R}$ whose union E is in $\mathcal{R}$. Then $\psi(\bar{E}) = \psi(\bar{E}) \cap \bar{E} = \bigcup_i (\psi(\bar{E}) \cap \bar{E}_i) \leq \bigcup_i \psi(\bar{E}_i)$ (for $\psi(\bar{E}) \varepsilon \theta(\mathcal{K})$). On the other hand, it is plain that $\bigcup_i \psi(\bar{E}_i) \leq \psi(\bar{E})$, so $\psi(\bar{E}) = \bigcup_i \psi(\bar{E}_i)$. Hence $r'(E) = r(\psi(E)) = \Sigma_i r(\psi(\bar{E}_i))$ (as $\psi(\bar{E}_i) \leq \bar{E}_i$, the $\psi(\bar{E}_i)$ are mutually disjoint) $= \Sigma_i r'(E_i)$. By the Radon-Nikodym theorem, $r'(E) = \int_E f(x)\,dr(x)$ for $E \varepsilon \mathcal{R}$, where f is a measurable function with values between 0 and 1. It is readily verified that $\psi(\psi(\bar{E})) = \psi(\bar{E})$, so $r(\psi(\psi(\bar{E}) = r(\psi(\bar{E}))$, whence $\int_{\psi(E)} f(x)\,dr(x) = r(\psi(E))$, where $\psi(E)$ is any set in the residue class $\psi(\bar{E})$. As $f(x) \geq 0$, it results that $f(x) = 1$ n. e. on $\psi(E)$, and as $r'(E - \psi(E)) = 0$, $f(x) = 0$ n. e. on $E - \psi(E)$. Now E is a general element of $\mathcal{R}$, so f is n. e. equal either to zero or to one, and hence is equivalent to the characteristic function of a measurable set K. It follows without difficulty that $\mathcal{K}$ is localized in K.

We close this section by giving a useful sufficient condition for a space to be localizable. For definitions and results concerning uniform spaces we refer to [9].

THEOREM 5. 2. *Any uniformly locally compact regular measure space is the direct sum of finite measure spaces (and hence localizable).*

Let $M = (R, \mathcal{R}, r)$ be a uniformly locally compact regular measure space, and let $N(.)$ be a uniform family of neighborhoods in R such that $N = N^{-1}$ and $N^2(x) \subset P(x)$ for $x \varepsilon R$, where $P(.)$ is a neighborhood family such that $P(x)$ is compact for $x \varepsilon R$. Then the sets $Q(x) = \bigcup_{i=1}^{\infty} N^i(x)$ are σ-compact, and open and closed, and it is easily seen that $Q(x)$ meets $Q(x')$ if and only if $Q(x) = Q(x')$. Hence R is the union of mutually disjoint σ-compact open and closed sets, say R_ξ, with ξ ranging over the index set Ξ.

Now any set in $\mathcal{R}$ is in the σ-ring determined by the compact subsets of R, and any such compact subset meets at most finitely many of the R_ξ.

Hence each element of $\mathcal{R}$ meets at most countably many of the R_ξ. It follows that M is the direct sum over Ξ of the measure spaces $M_\xi = (R_\xi, \mathcal{R}_\xi, r)$, where $\mathcal{R}_\xi = [E \cap R_\xi \mid E \in \mathcal{R}]$. Each M_ξ is clearly σ-compact and a regular locally compact measure space. Therefore M_ξ is σ-finite, and hence a direct sum of countably many finite measure spaces. It results that M is itself a direct sum of finite measure spaces.

COROLLARY 5. 2. *Any locally compact group is localizable, relative to Haar measure (on the conditional σ-ring determined by the compact subsets).*

Remark 5. 2. The following question arises fairly naturally in connection with this section: is any measure space strongly equivalent to a complete measure space (i. e. one in which every subset of a set equivalent to zero is also equivalent to zero)? Any measure space has a "completion" and it is easily seen that a localizable space is strongly equivalent to its completion, but the question appears to be open in general. An affirmative answer would be useful, but it seems plausible that the answer is negative.

We conclude this section by giving a partial extension of a theorem of Dieudonné [1], and also a short proof of the original theorem. As a matter of terminology, if $\mathfrak{N}$ is a subring of the measure ring of a measure space, a function on the space will be said to be measurable relative to $\mathfrak{N}$ if it is measurable in the usual sense relative to $\theta^{-1}(\mathfrak{N})$.

THEOREM 5. 3. *Let $M = (R, \mathcal{R}, r)$ be a localizable measure space with measure ring $\mathfrak{M}$ and let $\mathfrak{N}$ be a complete subring of $\mathfrak{M}$ containing $\bar{R}$. Let $\mathfrak{A}$ be the algebra of all bounded measurable functions on M that are also measurable relative to $\mathfrak{N}$. Then*

a) *there exists a localizable measure space $N = (S, \mathcal{S}, s)$ and an algebraic isomorphism ψ of $\mathfrak{A}$ onto the algebra of all bounded measurable functions on N (so that in particular $\mathfrak{N}$ is algebraically isomorphic to the measure ring of N);*

b) *there exists a positive continuous linear map ϕ of $L_1(M)$ to $L_1(N)$ such that for $f \in L_1(M)$ and $g \in \mathfrak{A}$,*

$$\int_M fg = \int_N \phi(f)\psi(g).$$

We mention that the measure space N is not unique within isomorphism or metric equivalence, tho necessarily unique within strong equivalence. In the special case (treated by Dieudonné) that M is finite there is a natural way of normalizing N, which has the useful feature that $\phi(f)$ is bounded when f is, but in general no such normalization appears possible.

The following lemma is somewhat more general than is strictly necessary for the proof of the theorem, but it has some independent interest.

LEMMA 6. 3. 1. *A complete subring of a complete measure ring is algebraically isomorphic to the measure ring of a measure space.*

If $(\mathfrak{M}, m)$ is a complete measure ring, then $\mathfrak{M}$ itself is isometrically algebraically isomorphic with the measure ring of a measure space. For the unit element I of $\mathfrak{M}$ is the l. u. b. of a collection of elements on which m is finite, from which it follows readily that it is the l. u. b. of a collection of mutually disjoint elements $\{A_\mu\}$ on each of which m is finite. The finite measure ring $(A_\mu \cap \mathfrak{M}, m)$ is well-known to be isometrically algebraically isomorphic with the measure ring of a measure space M_μ, and it follows without difficulty that $(\mathfrak{M}, m)$ is isometrically alegbraically isomorphic with the measure ring of the direct sum of the M_μ.

Now let $\mathfrak{N}$ be a complete subring of $\mathfrak{M}$, which we now take to be the measure ring of the measure space M. If I' is the l. u. b. of the elements of $\mathfrak{N}$, then $\mathfrak{N}$ is also a complete subring of the complete measure ring $(I' \cap \mathfrak{M}, m)$, so it is no essential loss of generality to assume that $I \varepsilon \mathfrak{N}$.

Let an element P of $\mathfrak{N}$ be called *finite* if there exists a nonzero element E in $\mathfrak{M}$ of finite measure such that when $A, B \varepsilon \mathfrak{N}$, $A, B \leq P$, and $A \cap E = B \cap E$, then $A = B$; we call E a *separating element* for P. Let $\mathfrak{F}$ be a family of nonzero finite elements of $\mathfrak{N}$ which is maximal with respect to mutual disjointness. Set $\text{l. u. b.}_{A \varepsilon \mathfrak{F}} A = Q$; we show next that $Q = I$. For otherwise there exists a nonzero element G of finite measure in $\mathfrak{M}$ such that $I - Q \geq G$. Putting R_0 for the l. u. b. of the elements R in $\mathfrak{N}$ for which $R \leq I - Q$ and $R \cap G = 0$, then $I - Q - R_0$ is disjoint from all the elements of $\mathfrak{F}$, is nonzero as $G \neq 0$, and is finite, since if $A, B \varepsilon \mathfrak{N}$, $A, B \leq I - Q - R_0$, and $A \cap G = B \cap G$, then $(A \ominus B) \cap G = 0$, while $A \ominus B \leq I - Q$, which by the definition of R_0 implies that $A \ominus B \leq R_0$. This inequality combined with the inequality $A, B \leq I - Q - R_0$, shows that $A \ominus B = 0$, or $A = B$.

Now put n for the function on $\mathfrak{N}$ defined by the equation

$$n(A) = \sum_{P \varepsilon \mathfrak{F}} m(\sigma(P) \cap A),$$

where $\sigma(P)$ is a separating element for P. Then it is clear that n is a countably additive non-negative-valued function, and to complete the proof it need only be shown that each element A of $\mathfrak{N}$ is the l. u. b. of elements on which n is finite. But it is clear that $[\sigma(P) \cap A \mid P \varepsilon \mathfrak{F}]$ is such a collection.

Proof of theorem. Let $N = (S, \mathcal{S}, s)$ be a (localizable) measure space whose measure ring is algebraically isomorphic to $\mathcal{N}$, and let $\mathcal{A}'$ be the algebra of all bounded measurable functions on N. Now every element in $\mathcal{A}'$ is the uniform limit of a sequence of finite linear combinations of characteristic functions of mutually disjoint measurable sets, and every such linear combination corresponds in an obvious way with a linear combination of characteristic functions of mutually disjoint elements of $\mathcal{N}$. As this correspondence is an isometry, it can be uniquely extended to an algebraic isomorphism between $\mathcal{A}'$ and the uniform closure of the set of all bounded finite linear combinations of characteristic functions of mutually disjoint elements of $\mathcal{N}$, which is simply $\mathcal{A}$. We put ψ for this isomorphism, as a map on $\mathcal{A}$ to $\mathcal{A}'$

Now let f be arbitrary in $L_1(M)$, and for an arbitrary measurable set E in N, put $\eta(E)$ for the element of $\mathcal{N}$ such that $\psi(\chi_{\eta(E)}) = \chi_E$. Then setting $m_f(E) = \int_{\eta(E)} f$, m_f is countably additive and is absolutely continuous with respect to s. As N is localizable, there exists a measurable function F on N such that $m_f(E) = \int_E F$ (altho m_f need not be bounded relative to s, the conclusion of the Radon-Nikodym theorem can be shown without difficulty to remain valid). Clearly F is non-negative when f is, and $\int_N F = \int_M f$, so F is integrable. Thus in case g is the characteristic function of a set in $\mathcal{N}$, we have $\int_M fg = \int_N \phi(f)\psi(g)$, if we set $F = \phi(f)$. It follows that this formula holds also for all finite linear combinations of such g's, hence for all finite linear combinations of characteristic functions of mutually disjoint sets in $\mathcal{N}$, and finally for all $g \varepsilon \mathcal{A}$.

THEOREM 5.4. (Dieudonné). *Let $M = (R, \mathcal{R}, r)$ be a finite perfect measure space and $\mathcal{T}$ a σ-subring of $\mathcal{R}$ containing R. Then there exists a perfect measure space $N = (S, \mathcal{S}, s)$, a continuous map ϕ of R onto S, and a positive linear map ψ of $\mathcal{B}(R)$ onto $\mathcal{B}(S)$ such that* a) *an element g of $\mathcal{B}(R)$ is (equivalent to) a function measurable relative to T if and only if it has the form $g = h \circ \phi$, for some $h \varepsilon \mathcal{B}(S)$, and for any E in $\mathcal{S}$,* $s(E) = \mathrm{r}(\phi^{-1}(E))$; b) *if $f \varepsilon \mathcal{B}(R)$ and $h \varepsilon C(S)$, then*

$$\int_M f(h \circ \phi) = \int_N \psi(f)h.$$

The "$\circ$" in the statement of the theorem signifies functional composition. We mention that if $m_y(y \varepsilon S)$ is the regular measure on R such that $\psi(f)(y) = \int f(x) dm_y(x)$ (by the Riesz-Markoff theorem, such a measure exists), then m_y is concentrated on $\phi^{-1}(y)$, as Dieudonné shows by a brief argument.

To prove the theorem, let $\mathcal{A}$ be the algebra of all continuous functions on R which are equivalent to functions measurable relative to $\mathcal{J}$. By well-known results, $\mathcal{A}$ is algebraically isomorphic to $\mathcal{B}(S)$, for some compact Hausdorff space S. Now $\mathcal{A}$ is a closed subring of $\mathcal{B}(R)$, so by a theorem of Stone [7], there exists a continuous map ϕ of R onto S such that $\mathcal{A}$ consists of all functions of the form $h \circ \phi$, with $h \varepsilon \mathcal{B}(S)$. Putting $s(E) = r(\phi^{-1}(E))$ for $E \varepsilon \mathcal{S}$, where $\mathcal{S}$ denotes the collection of all Borel sets in S, then $N = (S, \mathcal{S}, s)$ is a regular measure space, and can be shown without difficulty to be perfect (cf. e. g. the proof of Theorem 6. 1 below). If $f \varepsilon \mathcal{B}(M)$, and $m_f(E) = \int_{\phi^{-1}(E)} f \ (E \varepsilon \mathcal{S})$, then $| m_f | \leq \| f \| s(E)$, so by the Radon-Nikodym theorem there exists a bounded measurable function F such that $m_f(E) = \int_E F$, and as N is perfect, F can be taken to be in $\mathcal{B}(S)$. Putting $\psi(f) = F$, it follows by the argument used at the end of the proof of the last theorem that conclusion b) of the present theorem holds, and it has already been shown that a) holds.

6. Perfect spaces. In various situations it is desirable to replace a given measure space by a metrically equivalent one in which measurability difficulties are minimized, even tho the new space may, roughly sepaking, be much larger than the original one. In this section we show how a given space may be altered, or "perfected," in this way, and that this can be done in essentially only one way. The existence of a perfection of a finite measure space was shown by Kakutani, but his technique apparently does not extend to the general case. The uniqueness referred to in the following theorem is with respect to the following concept of identity of regular measure spaces: two such spaces are abstractly identical (as regular measure spaces) if they are isomorphic as measure spaces via an isomorphism which is also a topological homeomorphism.

THEOREM 6. 1. *Any measure space is metrically equivalent to a unique perfect space.*

Having shown that any measure space is metrically equivalent to a localizable space, it suffices to show that any localizable space is metrically equivalent to a unique perfect space. Let $M = (S, \mathcal{S}, s)$ be an arbitrary localizable measure space and let $\mathcal{A}_1$ be the algebra which results from adjoining the function which is identically one on S to the algebra $\mathcal{A}$ of all real-valued bounded measurable functions on M, which vanish at infinity. It is easy to see from well-known representation theorems that $\mathcal{A}_1$ is algebraically isomorphic to the algebra $\mathcal{R}(W)$ of all real-valued continuous functions on a compact Hausdorff space W. Now either $\mathcal{A} = \mathcal{A}_1$ or $\mathcal{A}$ is a maximal ideal in $\mathcal{A}_1$ (according as M is finite or not), and hence A corresponds either to $\mathcal{R}(W)$, or to a maximal ideal in $\mathcal{R}(W)$. Now every such ideal consists of all elements of $\mathcal{R}(W)$ which vanish at some fixed point w_0 of W. Putting $S' = W$, or $S' =$ the space obtained from W by deletion of w_0, according as $\mathcal{A} = \mathcal{A}_1$ or not, it follows that $\mathcal{A}$ is algebraically isomorphic with the algebra $\mathcal{R}(S')$ of all real-valued continuous functions which vanish at infinity on the locally compact Hausdorff space S'.

Let ϕ be the algebraic isomorphism from $\mathcal{A}$ onto $\mathcal{R}(S')$. Putting $\mathcal{A}_0$ for the set of elements of $\mathcal{A}$ which vanish outside sets of finite measure and $\mathcal{R}_0(S')$ for the subalgebra of $\mathcal{R}(S')$ consisting of elements which vanish outside of compact sets, then $\phi(\mathcal{A}_0) = \mathcal{R}_0(S')$, for $\mathcal{A}_0$ and $\mathcal{R}_0(S')$ may be characterized purely algebraically as the collections of elements p in the respective algebras for which there is an element q in the algebra such that $pq = p$. We now define a functional L on $\mathcal{R}_0(S')$ by the equation $L(f) = \int_S \phi^{-1}(f)(x)\,ds(x)$. It is easy to verify that L is linear, and non-negative on non-negative functions (note that ϕ is order-preserving, for it maps squares into squares). Hence there exists a regular measure space $M' = (S', \mathcal{S}', s')$, where $\mathcal{S}'$ is contained in the σ-ring determined by the compact subsets of S', such that $L(f) = \int_{S'} f(x)\,ds'(x)$ (see e. g. [2]). We show next that M' is perfect.

To show that the characteristic function of any chunk in M' coincides n. e. with an element of $\mathcal{R}(S')$, let E' be a chunk in M', and put A for the linear functional on $\mathcal{A}$ defined by the equation $A(f) = \int_{E'} \phi(f)\,ds'(x)$. It is easy to see that A is linear, and that for $f \varepsilon A_0$, $|A(f)| \leq \| f \|_1$. Since $\mathcal{A}_0$ is dense in $L_1(M)$, it results that the contraction of A to $\mathcal{A}_0$ can be extended to a bounded linear functional A_1 on $L_1(M)$. By Theorem 5.1 and the localizability of M there exists a bounded measurable function h on M such that $A_1(f) = \int_S f(x)h(x)\,ds(x)$, for $f \varepsilon L_1(M)$. It is plain that A is non-negative on non-negative functions, and hence that A_1 also is, from which it follows that h is n. e. non-negative-valued.

Now $h \,\varepsilon\, \mathcal{A}$, for let $F = [x \mid h(x) > \epsilon]$ where $\epsilon > 0$, and let $\{F_n; n = 1, 2, \cdots\}$ be an arbitrary monotone increasing sequence of chunks contained in F; to show that F is of finite measure, it suffices to show that the sequence $\{s(F_n)\}$ is bounded. Let f_n be the characteristic function of F_n, let $\{\alpha_n; n = 1, 2, \cdots\}$ be an arbitrary sequence of non-negative numbers for which the series $\sum_n \alpha_n$ is convergent, and set $f = \sum_n^{\infty} \alpha_n f_n$. Then it is easy to see that $f \,\varepsilon\, \mathcal{A}$, and for any positive integer n, $\sum_{1 \leq i \leq n} \alpha_i f_i \leq f$, which implies that $A(\sum_{1 \leq i \leq n} \alpha_i f_i) \leq A(f)$. It follows that $A(f) \geq \sum_{i=1}^{\infty} \alpha_i A(f_i) \geq \epsilon \sum_{i=1}^{\infty} \alpha_i s(F_i)$, so that the series $\sum_{i=1} \alpha_i s(F_i)$ is convergent. By virtue of the arbitrariness in the sequence $\{\alpha_i\}$, we can conclude that the sequence $\{s(F_i)\}$ is bounded. Hence, for all f in $\mathcal{A}_0$, $\int_{E'} \phi(f)(x') ds'(x) = \int_S f(x) h(x) ds(x) = \int_{S'} \phi(f)(x) \phi(h)(x) ds'(x)$, so that $\int_{S'} f'(x)(\chi_{E'}(x) - \phi(h)(x)) ds'(x) = 0$ for all $f' \,\varepsilon\, \mathcal{R}_0(S')$. Since $\mathcal{R}_0(S')$ is dense in $L_1(M')$, and since the function $k'(x) = \chi_{E'}(x) - \phi(h)(x)$ is bounded, it follows that $\int_{S'} f'(x) k'(x) ds'(x) = 0$ for $f' \,\varepsilon\, L_1(M')$. Hence $k'(x) = 0$ n. e., i. e. $\chi_{E'}(x) = \phi(h)(x)$ n. e.

To show that M' is perfect, it remains only to show that if the chunk E' in M' has an interior point, then $s'(E') > 0$. Now if E' has an interior point, there exists a non-negative continuous function f' on S' which is one at the interior point and which vanishes outside E' and has values between zero and one. Then $\chi_{E'} \geq f'$, so that $L(\chi_{E'}) \geq L(f')$, i. e.

$$s'(E') \geq \int_S \phi^{-1}(f')(x) ds(x),$$

and since $f' \geq 0$ and $f' \neq 0$, we have $\phi^{-1}(f') \geq 0$ and $\phi^{-1}(f) \neq 0$, and hence $\int_S \phi^{-1}(f')(x) ds(x) > 0$.

That M is metrically equivalent to M' now follows from Theorem 4. 3 combined with the following lemma (which is slightly more general than is necessary for the proof of the present theorem).

LEMMA 6. 1. 1. *For any bounded measurable function f on a perfect measure space there is a unique continuous function $\tilde{f}$ on the space which coincides n. e. with f. This function $\tilde{f}$ is necessarily bounded and vanishes at infinity as a function on a topological space if and only if f vanishes at infinity as a function on the measure space.*

It is readily seen to be sufficient to consider the case in which f is real-valued. Suppose to begin with that the perfect measure space $(T, \mathcal{J}, t)$ on which f is defined is such that T is compact. Putting

$$E_{i,n} = [x \mid i2^{-n} > f(x) \geq (i-1)2^{-n}] \quad (i = 0, \pm 1, \pm 2, \cdots; n = 1, 2, \cdots),$$

then $E_{i,n}$ is a chunk, and there exists an element $f_{i,n}$ of $\mathfrak{R}(T)$ which is equivalent to the characteristic function of $E_{i,n}$. Setting $f_n = \sum_i i2^{-n} f_{i,n}$, where the summation is over all i for which $f_{i,n} \neq 0$, the boundedness of f implies that this summation is finite. It results that $f_n \varepsilon \mathfrak{R}(T)$ and that $| f_n(x) - f(x) | \leqq 2^{-n}$ n. e. for $x \varepsilon T$, and finally that $| f_n(x) - f_{n+1}(x) | \leqq 2^{-n+1}$ n. e. for $x \varepsilon T$.

Now the complement in T of a set of measure zero is dense in T, for otherwise there would exist a non-void open subset of T of zero measure. Hence, $f_n - f_{n+1}$ being a continuous function in T and the inequality $| f_n(x) - f_{n+1}(x) | \leqq 2^{-n+1}$ holding n. e., this inequality holds for all $x \varepsilon T$. It follows that the series $\sum_{n=1}^{\infty} \{f_{n+1}(x) - f_n(x)\}$ is uniformly convergent, i. e. the sequence $\{f_n\}$ converges uniformly, say to $\tilde{f}$, where $\tilde{f}$ is evidently in $\mathfrak{R}(T)$. Plainly $\tilde{f}(x) = f(x)$ n. e. on T. Moreover, $\tilde{f}$ is the only continuous function on T which is n. e. equal to f, for the difference of two such functions is n. e. zero, and so by the remark at the beginning of the paragraph is identically zero.

Now let $N = (T, \mathcal{J}, t)$ be an arbitrary perfect measure space (so T need not be compact) and let f be an arbitrary bounded measurable function on N. If E is a chunk in N, there is a unique continuous function k on T which vanishes at infinity, and is equivalent to the characteristic function of E. It is plain that k^2 is equivalent to, and hence equal to, k. It results that k is the characteristic function of an open and closed subset E' of T, and since k vanishes at infinity on T, E' is compact. Now the contraction of f to E' is a bounded measurable function on the measure space $N_E = (E', \mathcal{J}_E, t)$, where $\mathcal{J}_E$ consists of the collection of all elements of T which are subsets of E. It follows from the preceding paragraph that there exists a continuous function f'_E on E', which is equivalent on N_E to the contraction of f (i. e. $f'_E(x) = f(x)$ n. e. for $x \varepsilon E'$).

Next we define $\tilde{f}$ on T as follows: if $x \varepsilon E'$, $\tilde{f}(x) = f'_E(x)$; $\tilde{f}$ is single-valued, for if $x \varepsilon E' \cap F'$, where F is a chunk in N, then f'_E and f'_F are both defined and continuous on $E' \cap F'$, and coincide n. e. on that set, and hence coincide everywhere there. It is clear that $\tilde{f}$ is continuous on T, and is bounded on any E' by the essential bound of f, and so is bounded on T. It is obvious that $\tilde{f}$ is the only continuous function on T which coincides n. e. with f. If moreover $\tilde{f}$ vanishes at infinity as a function on a topological space, then $[x \mid |f(x)| \geq e]$ is equivalent to the compact set $[x \mid |f(x)| \geq e]$, for any positive number e, so that f vanishes at infinity as a function on a measure space. Finally, if conversely f vanishes at infinity on N, then it is not difficult to see that $\tilde{f}$ is the uniform limit of a sequence $\{f_n\}$ of finite linear combinations of characteristic functions of chunks. Now f_n is in $\mathfrak{R}(T)$, and $|f_n(x) - f_m(x)|$ is equivalent on N to $|f_n(x) - f_m(x)|$, so that $\text{l. u. b.}_{x \varepsilon T} |f_n(x) - f_m(x)| = \text{Ess. l. u. b.}_{x \varepsilon T} |f_n(x) - f_m(x)|$. It results that the sequence $\{f_n\}$ converges uniformly. The limit is on the one hand equivalent to $\tilde{f}$ and on the other is a continuous function vanishing at infinity (being a uniform limit of continuous functions vanishing outside of compact sets). Thus f vanishes at infinity as a function on T.

It remains only to show that if two perfect spaces are metrically equivalent, then they are abstractly identical. Now if two measure spaces are metrically equivalent, the algebras of bounded measurable functions vanishing at infinity on them are algebraically isomorphic, by Theorem 4. 3. But in the case of a perfect space this algebra is by Lemma 6. 1. 1 naturally isometrically isomorphic to the algebra of all continuous functions on the space which vanish at infinity in the topological sense. Thus if $M = (R, \mathfrak{R}, r)$ and $M' = (R', \mathfrak{R}', r')$ are metrically equivalent perfect spaces, with the metric equivalence there is naturally associated an isometric isomorphism of the algebra of continuous functions vanishing topologically at infinity on R with the corresponding algebra on R'. By a theorem of Stone [7] (stated for compact spaces but extendable to locally compact spaces) this algebraic isomorphism can be implemented by a homomorphism of R with R', and there is no difficulty in showing that this homomorphism induces an isomorphism of M with M'.

The following theorem makes the results of the preceding section applicable to perfect spaces.

THEOREM 6. 2. *A perfect space is localizable.*

Let $N = (S, \mathcal{S}, s)$ be a perfect space, and let $\mathcal{K}$ be a conditional σ-ideal of chunks in N. By Theorem 5. 1, it suffices to show that $\mathcal{K}$ is localized. Each chunk in $\mathcal{K}$ is equivalent to an open and closed set in S, and the union of all such sets is an open set K. To show that $\mathcal{K}$ is localized in K it is necessary only to prove that if E is a chunk which is contained in K, then $\bar{E} \varepsilon \bar{\mathcal{K}}$. Since N is regular, there exists a sequence $\{C_n\}$ of compact subsets of E such that $\bar{E} \subset \bigcup_n \bar{C}_n$. Now $C_n \subset K$, so that C_n is covered by the open and closed sets equivalent to the elements of K, and hence is covered by a finite number of such open and closed sets. It results that $\bar{C}_n \,\varepsilon\, \bar{\mathcal{K}}$, and hence $\bar{E} \,\varepsilon\, \bar{\mathcal{K}}$.

UNIVERSITY OF CHICAGO.

REFERENCES.

[1] J. Dieudonné, "Sur le theoréme de Lebesgue-Nikodym (III)," *Annales de l'Université de Grenoble,* vol. 23 (1947-48), pp. 24-56, esp. pp. 39-42.

[2] P. R. Halmos, *Measure theory* (New York, 1950).

[3] S. Kakutani, "Concrete representation of abstract (L) spaces and the mean ergodic theorem," *Annals of Mathematics,* vol. 42 (1941), pp. 523-537.

[3a] D. Maharam, "On homogeneous measure algebras," *Proceedings of the National Academy of Sciences,* vol. 28 (1942), pp. 108-111.

[4] I. E. Segal, "Decomposition of operator algebras. I." To appear in *Memoirs of the American Mathematical Society.*

[5] ———, "Decompositions of operator algebras. II: Multiplicity theory." To appear in *Memoirs of the American Mathematical Society.*

[6] M. H. Stone, "Notes on integration, IV," *Proceedings of the National Academy of Sciences,* vol. 35 (1949), pp. 50-58.

[7] ———, "Applications of the theory of Boolean rings to general topology," *Transactions of the American Mathematical Society,* vol. 41 (1937), pp. 375-481.

[8] J. von Neumann, "On rings of operators. III," *Annals of Mathematics,* vol. 41 (1940), pp. 94-161, esp. pp. 127-129.

[9] A. Weil, "Sur les espaces a structure uniforme," *Actualités scientifiques et industrielles,* No. 551 (Paris. 1937).

UNE EXTENSION DU THÉORÈME DE VITALI.*

Par ARNAUD DENJOY.

1. Dans la théorie de la mesure euclidienne des ensembles appartenant à un espace cartésien, le théorème publié par Vitali en 1907[1] est parmi les plus importants. Ses applications aux propriétés des fonctions d'ensemble ou de variables réelles sont nombreuses. Par la promptitude, la simplicité des démonstrations utilisant ce théorème, celui-ci joue un role fondamental dans la théorie métrique des ensembles et des fonctions de nature cartésienne.

Il était naturel de chercher à étendre le théorème de Vitali, d'abord aux fonctions métriques quelconques pour les espaces cartésiens, ensuite aux métriques des espaces généraux.

Rappelons qu'une mesure euclidienne dans un espace cartésien U_r à r dimensions satisfait à cette condition que deux ensembles se déduisant l'un de l'autre par un déplacement euclidien, c'est-à-dire conservant les distances de deux points quelconques, ont même mesure, et que, dans le plan U_2, dans l'espace à trois dimensions U_3, etc., le carré, le cube, etc. de côté 1 ont l'aire, le volume, etc. égal à 1. Enfin, la mesure euclidienne utilisée par Vitali est une mesure borélienne, c'est-à-dire possédant l'additivité complète, et la soustractivité (voir au no. 6).

Généralement, une fonction d'ensemble, métrique $\phi(E)$, définie dans U_r est une quelconque fonction d'ensemble borélienne et non négative.

2. Toutes les tentatives d'élargir le champ de validité du théorème de Vitali ont sensiblement échoué. Nous allons en donner la raison.

En énonçant pour les espaces cartésiens, et plus précisément pour le plan U_2, les hypothèses de son théorème, c'est-à-dire des conditions suffisantes pour la validité de sa conclusion, Vitali a puisé trop largement dans les richesses de la nature métrique de ces espaces.

Dans l'espace U_r à r dimensions ($r > 2$), pour lequel le théorème de Vitali se formule immédiatement comme son auteur l'avait énoncé pour le

---

* Received October 25, 1950.

[1] Atti Accad. Sci. Torino, 43, 75-92 (1908). Voir aussi S. Banach (*Fundamenta Math.*, 5, 130-136 (1924). La présente étude, exposée pour la première fois à l'Université Johns Hopkins en septembre 1950, a fait l'objet de trois notes aux *Comptes Rendus de l'Ac. des Sc. de Paris* (t. 251, 1950, pp. 560-562; 600-601; 737-739.

plan U_2, dans U_r *à la mesure euclidienne d'ordre* r, qui est l'aire pour $r=2$, le volume pour $r=3$, et *qui seule figure dans la conclusion du théorème,* s'ajoutent une mesure pour les *distances* entre deux points et pour les longueurs des ensembles, une mesure des aires, etc. Mais puisque la conclusion du théorème de Vitali ne fait état, pour U_r, que de la mesure euclidienne d'ordre r, on conçoit qu'en introduisant dans les hypothèses, en plus de cette première mesure, de surcroît la notion de distance entre les points, on augmente considérablement la difficulté d'étendre le théorème à des espaces métrisés non cartésiens ou à des espaces cartésiens à métrique non euclidienne. Car en général aux métriques spéciales il ne correspond par un lien de caractère simple et nécessaire aucune mesure des distances entre les points.

Ce n'est pas tout. Les ensemble spécifiés dans les hypothèses de Vitali non seulement remplissent deux conditions de nature métrique, quant à leurs volumes d'une part (pour $r=3$), quant à leur dimension d'autre part (quelque soit r), mais il appartiennent encore à des espèces topologiques très particulières. Les uns sont des ensembles fermés, les autres ne diffèrent pas sensiblement d'ensembles ouverts. Or la conclusion du théorème de Vitali ne renferme aucune notion topologique.

3. Voici le théorème classique de Vitali pour l'espace à trois dimensions U_3:

Théorème de Vitali (énoncé A). *Si tout point M d'un ensemble H possédant un volume défini est centre d'une suite de parallélipipèdes rectangles $\omega_n=\omega_n(M)$, dont les côtés, tendant vers 0 quand n croît, ont leurs rapports mutuels compris entre deux nombres positifs k et $1/k$ $(0<k<1)$ indépendants de M et de n, et si chaque parallélipipède ω [identique à un $\omega_n(M)$] contient un ensemble* FERMÉ *γ dont le volume $V(\gamma)$ surpasse une fraction $\theta V(\omega)$ du volume de ω, θ étant indépendant de ω, il est possible de trouver parmi les γ une collection dénombrable d'ensembles $\gamma_1, \gamma_2\cdots, \gamma_i, \cdots$,* DISJOINTS, *et dont la réunion $\Gamma=\Sigma\gamma_i$ couvre H, à un ensemble de volume nul près. En outre on peut supposer que les points de Γ étrangers à H forment un ensemble de volume inférieur à un nombre positif indépendant ϵ préalablement donné.*

Etant d'abord bien spécifiée cette circonstance capitale que les γ_i sont *disjoints,* les conclusions du théorème tiennent en ces deux égalités, dont la première contient la partie essentielle de la proposition:

$$V(H-H\cdot\Gamma)=0, \qquad V(\Gamma-H\cdot\Gamma)<\epsilon.^{2}$$

---

[2] Je garde les signes $(+, \Sigma)$ pour signifier l'addition des ensembles et $(\cdot, \Pi)$ pour

Dans cette double conclusion il n'est question que de *volumes* d'ensembles. Cette dernière notion s'étend immédiatement aux fonctions métriques dans U_3, fonctions donnant par exemple à tout ensemble ouvert borné O une mesure $\phi(O)$ finie et non toujours nulle. Aux volumes des ensembles de U_3 on peut faire pareillement correspondre dans des espaces très généraux U une fonction d'ensemble $\phi(E)$ borélienne non négative. Un théorème généralisant pour l'espace U celui de Vitali pour U_3, concernera deux familles[3] d'ensembles associés chacun à chacun (ω, γ), la famille P groupant les $\omega = \omega(\gamma)$ et la famille G réunissant les $\gamma = \gamma(\omega)$; ω donné contient $\gamma(\omega)$, γ donné est dans $\omega(\gamma)$. Le rapport $\phi(\gamma)/\phi(\omega)$ devra être borné inférieurement, indépendamment du couple associé (ω, γ). L'ensemble H sera l'ensemble des points M communs à une suite d'ensembles ω *tendant vers* 0 *en un certain sens à préciser.* On devra assujettir ces familles P, G à des hypothèses telles qu'en conclusion, il existe dans G une collection d'ensembles *disjoints* $\gamma_1, \gamma_2, \cdots$, dont la réunion $\Gamma = \Sigma\gamma_i$ donne l'égalité $\phi(H - H \cdot \Gamma) = 0$.

Quant à la condition, moins remarquable *a priori,* $\phi(\Gamma - H \cdot \Gamma) < \epsilon$, il sera également possible de l'obtenir.

L'essentiel est que nous chercherons à exprimer les hypothèses en termes n'invoquant point l'existence dans l'espace U d'une seconde nature métrique différente de celle que manifeste la fonction $\phi(E)$.

Enfin nos hypothèses n'exigeront pas plus que nos conclusions l'existence d'une topologie de l'espace U.

4. Réduisons d'abord dans la mesure du possible le caractère particulier des hypothèses du théorème de Vitali en son énoncé A.

Soit O un ensemble ouvert contenant H. Nous pouvons retrancher de la famille des parallélipipèdes ω tous ceux qui, frontière comprise, ne sont pas totalement intérieurs à O, sans que la condition caractérisant les points M de

---

leur intersection. Je n'écris $E_1 - E_2$ que si l'ensemble E_1 contient l'ensemble E_2. Sinon, je note $E_1 - E_1 \cdot E_2$ l'ensemble des points de E_1 étrangers à E_2. La somme de E_1 et de E_2 décomposée en ensembles disjoints est $E_1 + (E_2 - E_1 \cdot E_2)$ ou $E_2 + (E_1 - E_1 \cdot E_2)$. Une expression telle que $E = (E_1 - E_1 \cdot E_2)(E_3 - E_3 \cdot E_4)(E_5 - E_5 \cdot E_6) + \cdot \cdot \cdot$ se développe selon les règles du calcul algébrique. M étant un point de l'espace contenant les ensembles, on retient les ensembles $E_i \cdot \cdot \cdot E_j$ contenant M. La somme de leurs coefficients est toujours 1 ou 0, 1 si M est dans, E, 0 si M est étranger à E.

Les notations fréquemment utilisées aujourd'hui $\cup$, $\cap$ et $E_1 - E_2$, même si E_1 ne contient pas E_2, ne se prêtent pas à l'application de règles aussi simples et précises.

[3] Nous distinguons d'une part les *familles* P, G d'ensembles ω, γ, P et G étant respectivement l'ensemble dont les ω, les γ constituent les éléments, d'autre part les *sommes* ou *réunions* de collections d'ensembles ω, γ. Celles-ci notées $\Sigma\omega$, $\Sigma\gamma$ sont les ensembles formés par les éléments appartenant aux divers ensembles ω, γ respectivement réunis.

H cesse d'être vérifiée par aucun de ces points. En effet, O étant ouvert, M situé dans O est à une distance positive 2δ de la frontière de O. A partir d'une certaine valeur de n, les côtés du parallélipipède $\omega_n(M)$ sont inférieurs à δ. Ces parallélipipèdes et leurs frontières sont dès lors intérieurs à O. Ils n'ont pas été compris dans les suppressions effectuées parmi les ω et qui, pour chaque point M ne font disparaitre qu'un nombre fini de parallélipipèdes $\omega_n(M)$.

Si H a un volume fini, et même si le volume de H est simplement déterminé à toute distance finie (mais infiniment grand avec cette distance), H peut être compris dans un ensemble ouvert O tel que $V(O-H) < \epsilon$. Dès lors, si l'on réduit les ω à ceux qui, avec leurs frontières, sont inclus dans O, tous les γ correspondants sont *a fortiori* inclus dans O, et il en est de même de tout ensemble ponctuel $\Gamma = \Sigma\gamma_i$ réunissant une collection d'ensembles γ. Mais Γ étant dans O, l'ensemble $\Gamma - \Gamma \cdot H$ des points de Γ étrangers à H est dans l'ensemble $O - H$ des points de O étrangers à H. Donc $V(\Gamma - \Gamma \cdot H) < \epsilon$.

Il suit de là que la démonstration de l'énoncé A du théorème de Vitali se réduit à celle du premier point, l'existence de la collection d'ensembles disjoints γ_i tels que, avec $\Gamma = \Sigma\gamma_i$, $V(H - H \cdot \Gamma) = 0$.

Enfin, si la démonstration est acquise pour le cas où $V(H)$ *est fini*, la conclusion vaudra aussi quand le volume de H, déterminé à toute distance finie (dans toute sphère ayant son centre à l'origine) ce volume devient infini avec cette distance. d désignant la distance d'un point de l'espace U_3 à l'origine, soit S_n la région intersphérique ouverte $(n-1 < d < n)$ et H_n l'ensemble $H \cdot S_n$, n prenant toutes les valeurs entières positives. $H - \Sigma H_n$ est formé de points situés sur les surfaces sphériques $d = n$, dont le volume total est nul. Donc $V(H - \Sigma H_n) = 0$. Soit ϵ_n un nombre positif terme d'une série dont la somme est inférieure à ϵ. Chacun des H_n ayant un volume fini, par hypothèse [la validité de l'énoncé A étant admise pour le cas de $V(H)$ fini], il existe pour chacun des H_n une famille $\Gamma_n = \sum\limits_i \gamma_i^n$ d'ensembles $\gamma_1^n, \gamma_2^n, \cdots$ disjoints et inclus dans S_n, vérifiant $V(H_n - H_n \cdot \Gamma_n) = 0$, et même en outre $V(\Gamma_n - H \cdot \Gamma_n) < \epsilon_n$; Γ_n et Γ_m sont disjoints si $m \neq n$. La collection $\Gamma = \Sigma\Gamma_n = \sum\limits_{i,n} \gamma_i^n$ est formée d'ensembles γ disjoints, et $V(H - H \cdot \Gamma) = 0$, $V(\Gamma - H \cdot \Gamma) < \epsilon$.

En résumé, le théorème de Vitali sous son énoncé A ne perd ni en généralité quand on suppose $V(H)$ fini, ni en portée si l'on néglige sa seconde conclusion $V(H - H \cdot \Gamma) < \epsilon$.

A ce théorème ainsi restreint et que nous appelons le *théorème A'*, nous allons substituer un autre énoncé B, renfermant moins de conditions de détail et parfaitement équivalent au premier.

5. Nous dirons qu'une famille P' d'ensembles ω' situés dans l'espace U *couvre indéfiniment au sens de la dimension* un point M, s'il existe une suite d'ensembles $\omega_1'(M), \omega'_2(M), \cdots$ appartenant à P', contenant M et dont les diamètres tendent vers 0.

THÉORÈME DE VITALI (énoncé B). *P' désignant une famille de sphères ouvertes ω' dont chacune contient un ensemble fermé $\gamma' = \gamma'(\omega')$ tel que $V(\gamma') > \theta' V(\omega')$ si $\gamma' = \gamma'(\omega')$, et le nombre positif θ' étant indépendant de ω', si l'ensemble H' indéfiniment couvert au sens de la dimension par la famille P' a un volume $V(H')$ déterminé, positif et fini, H' peut, à un ensemble de volume nul près, être couvert par une collection dénombrable $\gamma'_1, \gamma'_2, \cdots, \gamma'_i, \cdots$ d'ensembles γ'.*

Nous allons successivement supposer démontré le théorème sous sa forme A', puis sous sa forme B, et nous prouverons que la validité de la forme justifiée entraîne la validité de l'autre forme.

1° L'énoncé A' est supposé exact. Démontrons que B l'est aussi. Nous adoptons les hypothèses du théorème B. Englobons d'abord H' dans un ensemble ouvert O' de volume fini et dépassant si l'on veut de moins de ϵ le volume de H'. Nous ne changeons pas l'ensemble H' indéfiniment couvert par les ω' en supprimant d'entre ceux-ci toute sphère ω' telle que la sphère concentrique et de rayon 4 fois plus grand ne soit pas en totalité dans O'. La famille P' est réduite à une famille P'_0. Cela fait, si M' est un point de H' et si ω' est une sphère contenant M' et appartenant à P'_0, soit ω le cube de centre M' et de côtés (dirigés, si l'on veut, parallèlement aux axes de coordonnées) égaux au quadruple du rayon de ω'; ω contient ω'; ω est intérieur à O'. Au cube ω associons l'ensemble fermé $\gamma = \gamma(\omega)$ identique à $\gamma'(\omega')$; $\gamma(\omega)$ est dans ω, puisque $\gamma(\omega)$ est dans ω'. Le rapport $V(\gamma)/V(\omega)$ surpasse θ tel que θ/θ' soit égal au rapport de $V(\omega')$ à $V(\omega)$, soit $4\pi/3.64$; θ est indépendant de ω. Donc en vertu de l'énoncé A', H peut, à un ensemble de volume nul près, être couvert par une collection dénombrable de γ_i disjoints. Or ces γ_i sont des γ'_i de G'. L'énoncé B est donc une conséquence de l'énoncé A'.

2° L'énoncé B est supposé démontré. Reprenons les hypothèses de l'énoncé A' et montrons l'exactitude de sa conclusion. Si un point M de H est centre d'un parallélipipède rectangle ω de P, le plus grand côté l de celui-ci est tel que $l^3k^2 < V(\omega) < l^3$. Soit ω' la sphère de centre M et de rayon égal à $2l$; ω' contient ω. Nous pouvons avoir inclus préalablement H dans un ensemble ouvert 0 et avoir exclu de la famille des parallélipipèdes ω ceux dont le plus grand côté surpasse la moitié de la distance de M à la frontière de 0; ω' est dans O. A ω' nous faisons correspondre l'ensemble fermé $\gamma' = \gamma'(\omega')$

identique à $\gamma(\omega)$; γ' est dans ω' et $V(\gamma')/V(\omega') > \theta'$, si θ'/θ est le minimum du rapport $V(\omega)/V(\omega')$, soit $3k^2/32\pi$; θ' est indépendant de ω'. Soit H' l'ensemble indéfiniment couvert, au sens de la dimension, par la famille des ω'; H' contient H (et d'ailleurs H' est dans O dont le volume peut être supposé supérieur à celui de H de moins de ϵ)· En vertu du théorème B, nous pouvons couvir H', sauf éventuellement un ensemble de volume nul, par une collection d'ensembles disjoints $\gamma'_i = \gamma_i$. ·A fortiori H est-il couvert par les γ_i à un ensemble de volume nul près. L'énoncé A' est justifié.

En résumé l'énoncé B du théorème de Vitali équivaut entièrement à l'énoncé A. L'énoncé B réduit le théorème de Vitali à ses hypothèses essentielles, et c'est sous cette forme B que nous entendrons ci-après le théorème originel. Dans l'énoncé B donné plus haut nous prions le lecteur de supprimer les accents des lettres P', ω', γ', H', qui deviendront comme dans les énoncés A et A', et dans l'énoncé général ultérieur: P, ω, γ, H.

6. Dans le Théorème Général que nous voulons substituer au théorème de Vitali pour le rendre applicable à des espaces variés U, nos hypothèses ne feront appel ni à d'autre notion métrique pour U que celle de l'existence de la fonction $\phi(E)$, ni à une topologie de U. Dans certains cas, celui des espaces cartésiens par exemple, la fonction $\phi(E)$ se définit à partir de ses valeurs pour une famille d'ensembles élémentaires [dans U_3, ce sont les cubes précisés C_k, savoir: $p_i \leq 2^k x_i < p_i + 1$, $i = 1, 2, 3$; p_i entier quelconque]. La définition de $\phi(E)$ s'étend à une classe majeure d'ensembles qualifiés de mesurables-ϕ, par progrès borélien, c'est-à-dire par application de ces deux principes: Si les ensembles E_n sont disjoints et si tous les $\phi(E_n)$ existent, pour l'ensemble $E = \Sigma E_n$ le nombre $\phi(E)$ existe et vaut $\Sigma\phi(E_n)$. D'autre part, si $\phi(E)$ et $\phi(E')$ existent et si $E \supset E'$, $\phi(E - E')$ existe, et des lors vaut $\phi(E) - \phi(E')$. En conséquence si, le nombre $\phi(E_n)$ existant quel que soit n, l'ensemble E_n est soit constamment croissant, soit constamment décroissant, et si E est sa limite (au sens abstrait: $E = \Sigma E_n$ dans le premier cas, $E = \Pi E_n$ dans le second) $\phi(E)$ existe et vaut $\lim \phi(E_n)$, [toutefois quand E_n est décroissant, il faut supposer $\phi(E_n)$ fini, du moins à partir d'une certaine valeur de n].

La soustractivité d'une mesure borélienne s'entend ainsi au sens général. Quels que soient les ensembles E et E' non disjoints et tels que $\phi(E)$, $\phi(E')$ existent:

1° $\phi(E - E \cdot E')$ existe. En conséquence, d'après

$$E \cdot E' = E - (E - E \cdot E'),$$

$\phi(E \cdot E')$ existe, et aussi, de par $E + E' = E + (E' - E \cdot E')$, $\phi(E + E')$ existe;

2° $\phi(E - E \cdot E') = \phi(E) - \phi(E \cdot E')$.

Enfin, quel que soit l'ensemble E, il existe une mesure extérieure-ϕ notée $\phi_e(E)$ et qui est la borne inférieure stricte des mesures $\phi(E')$ des ensembles E' mesurables-ϕ et contenant E. La mesure intérieure $\phi_i(E)$ se définit de même par la borne supérieure stricte des mesures $\phi(E'')$ des ensembles E'' mesurables-ϕ intérieurs à E.

Les mesures extérieures et intérieurs $\phi_e(E)$ et $\phi_i(E)$ verifient une sorte d'additivité complète, se rattachant immédiatement à l'additivité complète des mesures-ϕ exactes· E étant la somme des ensembles E_n, si E_n est inclus dans un ensemble K_n mesurable-ϕ, et si les K_n sont disjoints, il en résulte $\phi_e(E) = \Sigma\phi_e(E_n)$; $\phi_i(E) = \Sigma\phi_i(E_n)$.

7. Pour aider le lecteur à pénêtrer pleinement le sens des hypothèses qui fonderont le Théorème Général, nous analysons ces conditions par avance.

Étant donnée une famille P d'ensembles ω non nécessairement mesurables-ϕ, nous dirons qu'*un point M de U est indéfiniment couvert, au sens de la métrique-ϕ*, par la famille P, s'il existe une suite d'ensembles ω_n contenant M et *dont la mesure extérieure $\phi_e(\omega_n)$ tend vers* 0 quand n croît.

Si dans U_3, $\phi(E)$ est le volume $V(E)$, un ensemble indéfiniment couvert par une famille d'ensembles ω au sens de la dimension l'est aussi au sens des volumes. Mais la réciproque n'est nullement vraie; $V(\omega)$ peut tendre vers 0 sans qu'il en soit ainsi du diamètre de ω. L'ensemble d'accumulation d'une suite d'ensembles ω_n, même ouverts, ayant un point commun et dont le volume tend vers 0 peut être identique à la totalité de l'espace U_3.

Donnons pour le plan U_2 et les aires $A(E)$ un exemple aisé à étendre à l'espace U_3 et aux volumes $V(E)$.

La région $u(\epsilon)$ comprise entre les courbes $y = -\epsilon/(1 + x^2)$ et $y = \epsilon/(1 + x^2)$ est finie et son aire est inférieure à 4ϵ. Cette région est coupée par la circonférence c ayant son centre à l'origine et le rayon 1 suivant un arc $a(\epsilon)$ sensiblement de longeur ϵ, d'ordonnées extrêmes $-\frac{1}{2}\epsilon$ et $\frac{1}{2}\epsilon$, d'abscisses voisines de 1. Soit $u'(\epsilon)$ la partie de $u(\epsilon)$ non intérieure à c. Sur c plaçons un ensemble parfait totalement discontinu dont les arcs contigus $a_1, a_2, \cdots, a_i, \cdots$ ont pour longueurs $\epsilon_1, \epsilon_2, \cdots, \epsilon_i, \cdots$; $\Sigma\epsilon_i \leq 2\pi$. En faisant tourner autour de l'origine $u'(\epsilon_i)$ ou plutôt $u'(\epsilon'_i)$, ϵ'_i différant peu de ϵ_i, nous appliquons l'arc $a(\epsilon'_i)$ sur l'arc a_i. A l'ensemble de régions ainsi obtenu ajoutons l'intérieur de c et les arcs a_i. Nous obtenons une région ω_0

d'aire inférieure à 9π; ω_0 évoque l'idée d'un poulpe dont l'intérieur de c formerait le corps et dont les régions infinies greffées sur les a_i seraient les tentacules, infinies en nombre et en dimension.

Dans ω_0 nous plaçons un ensemble fermé γ_0 obtenu en réduisant les rayons vecteurs r des points de ω_0 à $r' = r/(1 + \alpha r)$ $(0 < \alpha < 1$, α étant indépendant de $r)$, et en ajoutant à cette région sa frontière. $A(\gamma_0)/A(\omega_0)$ tend vers 1 quand α tend vers 0.

Si nous prenons les homothétiques de ω_0 et de γ_0 avec un rapport λ nous obtenons des ensembles $\omega_0(\lambda)$, $\gamma_0(\lambda)$. Soit $z_n = x_n + iy_n$ un ensemble partout dense (par exemple x_n et y_n prennent indépendamment l'un de l'autre toutes les valeurs réelles rationnelles). Soit λ_n tels que $\Sigma\lambda_n^2$ converge et soit inférieur à 1. Si θ est incommensurable avec π, imprimons dans U_2 figurant le plan complexe, à $\omega_0(\lambda_n)$ et à $\gamma_0(\lambda_n)$ la translation z_n et la rotation $n\theta$. La réunion de tous ces ensembles ouverts respectivement égaux aux $\omega_0(\lambda_n)$ est un ensemble ouvert ω_1 d'aire inférieure à 9π et la réunion des ensembles égaux aux $\gamma_0(\lambda_n)$ est un ensemble fermé à distance finie, que nous arrêtons à une distance assez grande de l'origine pour que l'ensemble fermé γ_1 restant, et qui est inclus dans ω_1, ait une aire supérieure à $\frac{1}{2}A(\omega_1)$.

Le complémentaire de ω_1 est un ensemble, fermé à distance finie, mais vraisemblablement totalement discontinu, bien que ω_1 ait une aire finie, inférieure à 9π. Des homothéties appliquées à ω_1 permettent de réduire cette aire au-dessous de toute limite. On conçoit qu'en constituant une famille P avec des ensembles ω se rattachant à ω_1 et à ses homothétiques, par reproduction complète ou par découpage arbitraire, on conçoit qu'avec ces ensembles ω dont chacun peut être partout dense dans le plan U_2, la notion d'ensemble indéfiniment couvert par ces ω, au sens des aires, s'éloigne de celle de l'ensemble indéfiniment recouvert au sens de la dimension.

Nous avons cherché à compliquer la tâche d'imaginer l'ensemble H indéfiniment couvert au sens des aires par une famille P d'ensembles ω. Mais il est facile d'obtenir un ensemble ouvert ω d'aire finie, arbitrairement petite, le complémentaire de ω étant totalement discontinu, ce qui implique évidemment la densité de ω dans tout le plan. Les deux familles de droites parallèles aux axes Ox, Oy et dont la coordonnée constante est rationnelle forment un ensemble d'aire nulle. On peut donc inclure celui-ci dans un ensemble ouvert ω d'aire inférieure à ϵ. Le complémentaire de ω est évidemment totalement discontinu.

8. Revenons à nos hypothèses sur les familles ω, γ. Nous avons dit que l'ensemble H et les ensembles ω désignés dans l'un ou l'autre des énoncés

A et B du théorème de Vitali remplissent, au sens des volumes jouant le rôle de fonction métrique-ϕ, la condition de recouvrement indéfini.

Seulement dans ce théorème de Vitali, les ω qui sont des parallélipipèdes, comprenant ou non leurs frontières, et les γ qui sont des ensembles fermés, sont à la fois euclidiennement mesurables en volumes, et en outre doués de caractères topologiques remarquables.

Les ensembles ω de notre énoncé ne seront pas nécessairement mesurables-ϕ, leur mesure extérieure intervenant seule.[4] Les γ au contraire *seront nécessairement mesurables-ϕ, et de mesure positive* mais ni les ω ni les γ n'ont à remplir d'autre condition de nature, *l'espace U n'étant pas obligatoirement doué d'une topologie.*

9. Voici les conditions que nous allons imposer aux systèmes (ω, γ) pour assurer la validité du théorème général que nous avons en vue.

1° Quel que soit l'ensemble γ de la famille G, les points étrangers à γ et indéfiniment couverts au sens de la métrique-ϕ par les ω' joints à γ forment un ensemble $\rho(\gamma)$ de mesure-ϕ égale à zéro.

Supposons choisi l'ensemble γ_1 désigné dans la conclusion du théorème à énoncer. Il s'agira de trouver les autres γ_i $(i \geqq 2)$. Avec la simple hypothèse $\gamma_i \subset \omega(\gamma_i)$, nous ne sommes assurés que les γ_i sont disjoints de γ_1 que si les $\omega(\gamma_i)$ sont eux-mêmes disjoints de γ_1. Donc ces $\omega(\gamma_i)$ seront dans la famille P_1 des ensembles ω restant de la famille P quand on retranche de celle-ci les ω joints à γ_1. *A priori* ces ω joints à γ_1 peuvent couvrir indéfiniment outre les points de $H \cdot \gamma_1$ ceux d'un ensemble $\rho(\gamma_1)$ appartenant à H. Il est possible que $\rho(\gamma_1)$ soit totalement disjoint des ω formant P_1 et dans ce cas la collection Γ des ensembles γ_i $(i \geqq 1)$ ne contiendrait aucun point de $\rho(\gamma_1)$. La conclusion du théorème ne peut donc pas être générale si $\rho(\gamma_1)$ a une mesure extérieure-ϕ positive. Et c'est ainsi que la nécessité de notre première condition apparaît.

Evidemment $\rho(\gamma)$, s'il existe, est dans H, puisque tout point de $\rho(\gamma)$ est indéfiniment couvert (au sens de la métrique-ϕ) par une famille partielle ω' incluse dans P.

Avec les hypothèses de l'énoncé B de Vitali, γ est un ensemble fermé contenu dans une sphère $\omega(\gamma)$. Les sphères ω' jointes à γ et dont le volume tend vers 0 ont évidemment leurs points d'accumulation dans γ. Elles ne peuvent donc couvrir indéfiniment aucun point étranger à γ; $\rho(\gamma)$ n'existe pas dans l'hypothèse de Vitali. La condition $V[\rho(\gamma)] = 0$ est verifiée.

---

[4] Quand l'espace U est cartésien, la considération d'ensembles ω non mesurables (nonboréliens) n'est pas d'une très grande utilité. Une des idées les plus importantes dues à Lebesgue, est que pratiquement l'Analyse ne peut conduire qu'à des ensembles mesurables, puisque les passages à la limite conservent ce caractère. On peut toutefois songer à des sommes non dénombrables d'ensembles mesurables, sommes qui eventuellement ne seraient pas mesurables.

Mais si, dans U_3 et avec la métrique des volumes euclidiens, chacun des ω peut être partout dense, on conçoit que cette première hypothèse introduit une restriction effective. La démonstration en fera connaître le rôle essentiel.

Soit M un point de U étranger à $\gamma + \rho(\gamma)$. Les ω' contenant M et joints à γ ont leur mesure extérieure-ϕ bornée inférieurement par un nombre β. Il n'y a pas d'ensemble ω' joint à γ et contenant M, et pour lequel $\phi_e(\omega') < \beta$. Si M est en même temps dans H, il y une infinité d'ensembles ω'' contenant M et tels que $\phi(\omega'')$ tende vers 0. Mais ces ω'' sont tous disjoints de γ dès que $\phi(\omega'') < \beta$.

2° Quel que soit γ dans G, il existe deux nombres a et b, vérifiant $1 < a < b$, indépendants de γ, et tels que les $\omega'(\gamma')$ joints à γ et vérifiant la condition $\phi(\gamma') < a\phi(\gamma)$ forment par leur réunion un ensemble $\Omega(\gamma)$ dont la mesure extérieure-ϕ est inférieure à $b\phi(\gamma)$.

$\omega(\gamma)$ est évidemment un des ω' puisque $\omega(\gamma)$ est joint à γ qu'il contient et que $\phi(\gamma) < a\phi(\gamma)$, d'après $a > 1$. Donc $\omega(\gamma) \subset \Omega(\gamma)$ et $\phi_e[\omega(\gamma)] < b\phi(\gamma)$.

En conséquence, si M est étranger à $\gamma + \rho(\gamma)$, les γ' tels que $\omega(\gamma')$ soit joint à γ et contienne M ont leur mesure $\phi(\gamma')$ supérieure à un nombre positif α, dépendant évidemment de M et de γ. Car d'une inégalité $\phi_e[\omega(\gamma')] > \beta$ résulte $\phi(\gamma') > \beta/b$, ce dernier nombre pouvant tenir le rôle de α.

Sous la forme B du théorème de Vitali, il est aisé de voir que la seconde de nos hypothèses est vérifiée quel que soit $a > 1$. De plus, tandis que la connaissance des nombres a et b, de $\phi(\gamma)$ et d'un point N de γ ne suffit pas à définir dans U un ensemble contenant certainement $\Omega(\gamma)$ indépendamment de la donnée de la famille P des ω, puisque ceux-ci peuvent être partout denses dans un espace U, au contraire dans les conditions du théorème B de Vitali, la connaissance de a, du coefficient θ, du volume $V(\gamma)$ et d'un point de γ permet de définir une sphère ayant ce dernier point pour centre et contenant $\Omega(\gamma)$, indépendamment de la détermination particulière de la famille P des sphères ω considérées dans l'énoncé B de Vitali.

En effet, d'après l'hypothèse $V(\gamma) > \theta V(\omega)$, ω étant $\omega(\gamma)$, la sphère ω a un rayon inférieur à $r = \{3V(\gamma)/(4\pi\theta)\}^{1/3}$. Si le point N est dans γ, la totalité de γ est dans la sphère de centre N et de rayon $2r$. Le nombre a étant quelconque supérieur à 1, si $V(\gamma') < aV(\gamma)$, $\omega(\gamma')$ a un rayon inférieur à $r' = a^{1/3}r$, et si en outre la sphère $\omega' = \omega(\gamma')$ est jointe à γ, elle est entièrement comprise dans la sphère de centre N et de rayon $2(1+a^{1/3})r$. Cette sphère contient l'ensemble $\Omega(\gamma)=\Sigma\omega'$. Son volume est inférieur au produit de $V(\gamma)$ par un nombre b fonction de a et de θ, indépendant de la composition des familles P et G. Cette sphère contenant $\Omega(\gamma)$ est déterminée par la connaissance des seuls nombres a, θ, $V(\gamma)$ et du point N de γ, centre de cette sphère.

3° La mesure extérieure-ϕ de l'ensemble D réunissant les ensembles γ est finie.

Nous montrerons plus loin (16) la nécessité de cette troisième hypothèse pour la validité du théorème général.

Dans le cas due théorème de Vitali sous sa forme B, l'ensemble indéfiniment couvert H a un volume fini et tout ensemble ouvert O contenant H, O ayant par exemple un volume excédant celui de H de moins de ϵ, peut être considéré comme contenant tous les ensembles γ.

Les trois hypothèses que nous retenons sont donc satisfaites dans le cas particulier régi par le théorème de Vitali.

10. Voici maintenant notre énoncé:

Théorème Général. *Dans l'espace U où une fonction d'ensemble $\phi(E)$ non négative, borélienne, est définie et sert de mesure, on suppose données deux familles d'ensembles (ω,γ) associés, P famille d'ensembles $\omega=\omega(\gamma)$ non nécessairement mesurables-ϕ, G famille d'ensembles $\gamma=\gamma(\omega)$, $\gamma(\omega)$ étant inclus dans ω et possédant une mesure $\phi(\gamma)$ déterminée, positive et finie,* et ces familles vérifiant les conditoins suivantes:

1° *Quel que soit l'ensemble γ dans G, les points étrangers à γ et indéfiniment couverts au sens de la métrique-ϕ par les ω' de P joints à γ forment un ensemble $\rho(\gamma)$ dont la mesure-ϕ est nulle*;

2° *il existe deux nombres positifs a et b $(1<a<b)$ indépendants de γ quelconque dans G et tels que l'ensemble $\Omega(\gamma)$ réunissant les ω' joints à γ et vérifiant $\phi(\gamma')<a\phi(\gamma)$ si $\gamma'=\gamma'(\omega')$ a une mesure extérieure-ϕ inférieure à $b\phi(\gamma)$.*

3° *La mesure extérieure-ϕ de l'ensemble D réunissant les γ est finie;*

sous ces conditions suffisantes, *H désignant l'ensemble des point M de U indéfiniment couverts au sens de la métrique-ϕ par la famille P des ensembles ω* [M appartient à une suite d'ensembles ω_n de P dont la mesure extérieure $\phi_e(\omega_n)$ tend vers 0], *et Δ, inclus dans H, étant l'ensemble des points indéfiniment couverts au sens de la métrique-ϕ par la famille G des γ:*

1° *L'ensemble H est couvert, à un ensemble de mesure-ϕ nulle près,* *par une collection dénombrable d'ensembles* $\gamma_1, \gamma_2, \cdots$ *de* G, *disjoints. Si* $\Gamma = \Sigma\gamma_i$, $\phi(H - H \cdot \Gamma) = 0$;

2° H *et* Δ *sont mesurables-*ϕ, *l'excès de* H *sur* Δ *ayant la mesure-*ϕ *nulle* $[\phi(H) = \phi(\Delta)]$. *De plus, si petit que soit le nombre positif* ϵ, *on peut choisir les* γ_i *de* G *de façon que* $\phi(\Delta) \leqq \phi(\Gamma) < \phi(\Delta) + \epsilon$.

11. Pour démontrer la première conclusion de ce théorème, nous suivrons la marche indiquée par Banach pour établir le théorème classique de Vitali.

De $D = \Sigma\gamma$ résulte $\phi(\gamma) < \phi_e(D)$, quel que soit γ dans G; d'après la troisième hypothèse: $[\phi_e(D) < \infty]$, les nombres $\phi(\gamma)$ sont bornés supérieurement. Soit μ_1 leur maximum (borne supérieure stricte) et γ_1 un ensemble γ de G tel que $\phi(\gamma_1) > \mu_1/a$; γ_1 existe d'après $a > 1$. Quel que soit γ dans G, $\phi(\gamma) \leqq \mu_1 < a\phi(\gamma_1)$.

Il peut se faire que précisément les points de H étrangers à γ_1 forment un ensemble de mesure-ϕ nulle. Si $\phi(H - H \cdot \gamma_1) = 0$, la proposition est vérifiée. Supposons qu'il en soit autrement.

Posons $\rho_1 = \rho(\gamma_1)$; ρ_1 est dans H, et $H \cdot \rho_1 = \rho_1$; d'après la première hypothèse, $\phi(\rho_1) = 0$. Soit $H_1 = H - H(\gamma_1 + \rho_1)$, ou $H_1 = H - H \cdot \gamma_1 - \rho_1$; H_1 et $H - H \cdot \gamma_1$ ont la même mesure extérieure, et nous supposons que celle-ci est positive. Si M est étranger à $\rho_1 + \gamma_1$, il existe un nombre α_1, dépendant de M (et de γ_1 évidemment) tel que si $\phi(\gamma) < \alpha_1$ et si $\omega(\gamma)$ contient M, $\omega(\gamma)$ n'est pas joint à γ_1. Soit M un point de H_1; H_1 étant dans H, M appartient à H, donc M est contenu dans une suite d'ensembles ω de mesure extérieure tendant vers 0. Et M inclus dans H_1 est étranger à $\gamma_1 + \rho_1$. Donc dès que, ω contenant M, $\phi_e(\omega)$ est assez petit pour que $\phi(\gamma) < \alpha_1$, γ étant $\gamma(\omega)$, toute la suite restante des ω est disjointe de γ.

De P retranchons tous les ω joints à γ_1 et soit P_1 la famille restante. Nous venons de montrer qu'au sens de la métrique-ϕ, la famille P_1 couvre indéfiniment tout point de H_1. Soit H'_1 l'ensemble indéfiniment couvert, au sens de la métrique-ϕ, par la famille P_1; H'_1 contient H_1, nous venons de le dire. D'ailleurs H'_1 est évidemment inclus dans H, de même que la famille P_1 est incluse dans la famille P; tout point indéfiniment couvert par P_1 l'est *a fortiori* par P; Mais comme P_1 ne contient aucun ω joint à γ_1, H'_1 ne renferme aucun point de γ_1; H'_1 inclus dans H est disjoint de $H \cdot \gamma_1$; H'_1 contenant $H_1 = H - H \cdot \gamma - \rho_1$ est contenu dans $H - H \cdot \gamma_1 = H_1 + \rho_1$. ($H_1$ est disjoint de ρ_1, puisque H_1 se déduit de H par extraction non seulement des points appartenant à γ_1, mais encore de tous les points de ρ_1, ceux-ci étant étrangers à γ_1). Les points de H'_1 étrangers à H_1 sont donc dans ρ_1. Si

$H'_1 = H_1 + \rho'_1$, ρ'_1, comme ρ_1 dont il est une partie, a la mesure-ϕ nulle. D'ailleurs ρ'_1 peut exister, H'_1 peut contenir des points de ρ_1. Il est évidemment possible qu'un point étranger à γ_1 et couvert indéfiniment (au sens de la métrique-ϕ) par les ω de P joints à γ_1 soit aussi couvert indéfiniment par les ω de P disjoints de γ_1, et qui forment P_1.

Soit G_1 l'ensemble des $\gamma = \gamma(\omega)$ correspondant aux ω de P_1. Puisque tout γ de G_1 est dans un ω disjoint de γ_1, *a fortiori* tous les γ de G_1 sont disjoints de γ_1. D'autre part le maximum μ_2 de $\phi(\gamma)$ pour les γ de G_1 vaut au plus μ_1 et il est inférieur à $a\phi(\gamma_1)$.

12. Maintenant, avec les familles P_1 et G_1 nous procédons comme nous l'avons fait avec P et G. Cette règle engendre une suite d'opérations qui nous donnent une succession d'ensembles $\gamma_1, \gamma_2, \cdots, \gamma_i, \cdots$ disjoints, se déterminant (même à partir de γ_1, si $P = P_0$, $G = G_0$) de la façon suivante:

Soit $\Gamma_n = \gamma_1 + \gamma_2 + \cdots + \gamma_n$. Si $H - H \cdot \Gamma_n$ a la mesure-ϕ nulle, la proposition est vérifiée. Supposons que cet ensemble ait une mesure extérieure-ϕ positive. Posons $\rho_i = \rho(\gamma_i)$. D'après la première hypothèse, l'ensemble $\rho_1 + \rho_2 + \cdots + \rho_n = \Theta_n$ a la mesure-ϕ nulle. Il est inclus dans H. Mais Θ_n n'est pas nécessairement disjoint de Γ_n. Car ρ_i, disjoint de γ_i, ne l'est pas nécessairement de tous les γ_j pour $j \neq i$. Soit $H_n = H - H \cdot (\Gamma_n + \Theta_n)$. Par hypothèse $\phi_e(H_n) > 0$.

Soit M un point quelconque de H_n; M étant dans H, il existe une suite d'ensembles ω contenant M et tels que $\phi_e(\omega)$ tende vers 0. D'autre part, M étant étranger à $\Gamma_n + \Theta_n$ est étranger à tout ensemble $\gamma_i + \rho_i$ pour $1 \leq i \leq n$. Donc il existe un nombre α_i dépendant de M et de γ_i, tel que si $\phi(\gamma) < \alpha_i$ et si $\omega = \omega(\gamma)$ contient M, $\omega(\gamma)$ est disjoint de γ_i.

Soit α'_n est le plus petit des nombres $\alpha_1, \alpha_2, \cdots, \alpha_n$. Considérons les ω vérifiant ces conditions: ω contient M et $\phi(\gamma) < \alpha'_n$ si $\gamma = \gamma(\omega)$. Ces ω sont tous disjoints de Γ_n et ils renferment parmi eux une suite dont la mesure extérieure-ϕ tend vers 0. En conséquence, si l'on retranche de P tous les ω joints à Γ_n, c'est-à-dire à l'un au moins des γ_i pour $i = 1, 2, \cdots, n$, la famille P_n des ω restants couvre indéfiniment, au sens de la métrique-ϕ, l'ensemble H_n. D'ailleurs, l'ensemble H'_n indéfiniment couvert par les P_n et qui est inclus dans H ne contient aucun point de Γ_n, puisque les ω de P_n sont tous disjoints de Γ_n. Donc H'_n est contenu dans $H - H \cdot \Gamma_n$. En conséquence H'_n contient H_n et est inclus dans $H - H \cdot \Gamma_n$; $H - H \cdot \Gamma_n$ excède H_n d'un ensemble inclus dans ou identique à Θ_n. Donc $H'_n = H_n + \Theta'_n$, Θ'_n étant une partie de Θ_n; Θ'_n peut exister, un même point pouvant être couvert indéfiniment (au sens de la métrique-ϕ) à la fois par la famille des ω de P joints à Γ_n et par la famille des ω de P disjoints de Γ_n et qui forment P_n. En tous cas $\phi(\Theta'_n) = \phi(\Theta_n) = 0$.

Soit G_n l'ensemble des γ correspondant aux ω de P_n. Si γ est dans G_n, $\omega(\gamma)$ est disjoint de Γ_n. Comme γ est inclus dans $\omega(\gamma)$, *a fortiori* γ est disjoint de Γ_n, c'est-à-dire de tous les γ_i pour $i \leqq n$.

Nous désignons par μ_{n+1} le maximum de $\phi(\gamma)$ pour les γ de G_n. Evidemment $\mu_{n+1} \leqq \mu_n \leqq \phi_e(D)$. Soit γ_{n+1} un des γ de G pour lequel $\phi(\gamma_{n+1}) > \mu_{n+1}/a$. Quel que soit γ dans G_n, $\phi(\gamma) < a\phi(\gamma_{n+1})$; γ_{n+1} étant dans G_n, est disjoint des γ_i antérieurs $(i \leqq n)$.

Ayant défini γ_{n+1}, nous sommes maintenant passés de Γ_n à Γ_{n+1}. Le mécanisme de la détermination progressive des γ_i est entièrement établi. Ces γ_i seront tous disjoints les uns des autres.

Les familles G_n décroissent, G_n contenant la totalité de G_{n+1} et en outre tout au moins l'ensemble γ_{n+1}.

13. Il reste à examiner le cas où la suite γ_i est illimitée, c'est-à-dire le cas où nous ne sommes pas arrêtés par une suite finie $\gamma_1, \gamma_2, \cdots, \gamma_n$ telle que $\phi(H - H \cdot \Gamma_m) = 0$.

Soit $\Gamma = \Sigma\gamma_i$. Nous voulons montrer l'égalité $\phi(H - H \cdot \Gamma) = 0$. Soit $\Theta = \Sigma\rho_i$; $\phi(\Theta) = 0$, puisque chacun des $\phi(\rho_i)$ est nul. Θ est dans H; $H - H \cdot \Gamma = R$ et $H - H \cdot (\Gamma + \Theta) = R'$ ont la même mesure extérieure. L'égalité $\phi(R) = 0$ équivaut à $\phi(R') = 0$. Prouvons celle-ci.

Observons que, les γ_i étant mesurables-ϕ et disjoints, Γ est mesurable-ϕ et $\phi(\Gamma) = \Sigma\phi(\gamma_i)$. L'inclusion $\Gamma \subset D$ résultant du fait que tout γ_i est dans D, entraîne $\phi(\Gamma) \leqq \phi_e(D) < \infty$. Par suite, la série $\Sigma\phi(\gamma_i)$ converge et, en conséquence $\phi(\gamma_i)$ tend vers zéro, quand i croît.

Soit M un point de R'. Donnons-nous un nombre positif η quelconque et soit N un entier tel que $\sum_{n>N} \phi(\gamma_n) < \eta$.

M étant dans R' est inclus dans H et étranger à $\Sigma(\gamma_i + \rho_i)$. Donc M est étranger à $\gamma_i + \rho_i$ quel que soit i. Soit α (ou α'_N) le plus petit des nombres $\alpha_1, \cdots, \alpha_N$; M étant inclus dans H, il y a une infinité d'ensembles ω de mesure extérieure tendant vers 0 et contenant M. Pour les γ correspondants, $\phi(\gamma) < \phi_e(\omega)$ et $\phi(\gamma)$ tend aussi vers 0. Considérons un ω pour lequel $\gamma = \gamma(\omega)$ vérifie $\phi(\gamma) < \alpha$; ω est disjoint de $\gamma_1, \cdots, \gamma_N$; ω est donc dans P_N, et γ est dans G_N. Mais $\phi(\gamma_n)$ tendant vers 0 quand n croît, il y a un premier m tel que $a\phi(\gamma_{m+1}) \leqq \phi(\gamma)$, tandis que $\phi(\gamma) < a\phi(\gamma_n)$ pour $n \leqq m$; γ n'est donc pas dans G_m, et $m \geqq N + 1$. Donce ω n'est pas dan P_m.

Les P_n, les G_n décroissent quand n croît. Il y a donc un i vérifiant $N+1 \leq i \leq m$, tel que γ est dans G_{i-1} (ensemble des γ tels que $\omega(\gamma)$ est disjoint de Γ_{i-1}) sans être dans G_i (ensemble des γ disjoints de Γ_i). Donc $\omega(\gamma)$ est joint à γ_i et, d'après $\phi(\gamma) < a\phi(\gamma_i)$; $(i \leq m)$, $\omega(\gamma)$ est dans l'ensemble $\Omega(\gamma_i)$, tout en contenant M. Donc M est dans $\Omega(\gamma_i)$. Soit $\Omega'_N = \sum_{n>N} \Omega(\gamma_n)$; M est dans Ω'_N; M étant un point quelconque de R', il s'ensuit $R' \subset \Omega'_N$. Mais, en vertu de la deuxième hypothèse, $\phi_e[\Omega(\gamma_n)] < b\phi(\gamma_n)$. Donc

$$\phi_e[\Omega'_N] < b \sum_{n>N} \phi(\gamma_n) < b\eta.$$

Donc $\phi(R') < b\eta$ quel que soit η positif; $\phi(R') = 0$ et $\phi(R) = 0$. La première conclusion du théorème est établie.

14. Passons à la seconde conclusion. Elle se décompose en deux parties. Montrons d'abord que H et Δ sont mesurables-ϕ.

Une conséquence évidente de la première conclusion du Théorème est que, Γ étant dans $D = \Sigma\gamma$, il en est de même de $H \cdot \Gamma$. Donc la partie de H étrangère à D est *a fortiori* étrangère à Γ. Elle est dans $R = H - H \cdot \Gamma$. C'est un ensemble de mesure-ϕ nulle.

η_n étant un nombre positif tendant vers 0 en décroissant quand n croît, soit D_n la réunion des γ vérifiant $\phi(\gamma) < \eta_n$. Si G^n est leur famille et si P^n est la famille des ω correspondant, P^n, qui comprend tous les ω pour lesquels $\phi_e(\omega) < \eta_n$, couvre indéfiniment, au sens de la métrique-ϕ, le même ensemble H que P elle-même. Les ensembles D_n, les familles G^n, P^n sont non croissants quand n croît.

D'après la première conclusion établie, on peut couvrir H, exception faite d'un ensemble R_n de mesure-ϕ nulle, avec une suite d'ensembles $\gamma_1^n, \gamma_2^n, \cdots, \gamma_i^n, \cdots$ disjoints, appartenant à la famille G^n, donc situés dans D_n. Si $\Gamma_n = \sum_i \gamma_i^n$, l'ensemble $R_n = H - H \cdot \Gamma_n$ vérifie $\phi(R_n) = 0$.

Soit Δ l'ensemble ΠD_n, commun à tous les D_n; Δ est l'ensemble indéfiniment couvert par la famille G, au sens de la métrique-ϕ. Car tout point indéfiniment couvert par la famille G au sens de la métrique-ϕ appartient à une suite de γ tels que $\phi(\gamma)$ tende vers 0. Ce point appartient donc à tous les D_n. Il est donc dans Δ. Réciproquement, tout point de Δ, étant commun à tous les D_n, appartient à un γ pour lequel $\phi(\gamma) < \eta_n$, quel que soit n. Ce point est couvert indéfiniment par la famille G, au sens de la métrique-ϕ.

Evidemment $\Delta \subset H$, puisque un point couvert indéfiniment (au sens de la métrique-ϕ) par les γ l'est *a fortiori* par les $\omega(\gamma)$ correspondants. En effet, d'une part $\omega(\gamma)$ contient γ, d'autre part d'après $\phi_e[\omega(\gamma)] < b\phi(\gamma)$ (conséquence de la seconde hypothèse), $\phi_e[\omega(\gamma)]$ tend vers 0 quand il en est ainsi de $\phi(\gamma)$.

Soit $\Delta' = \Pi\Gamma_n$. Tous les Γ_n sont mesurables-ϕ. Donc Δ' est également mesurable-ϕ; Δ' est inclus dans Δ. Car, Γ_n étant dans D_n, $\Pi\Gamma_n = \Delta'$ est dans $\Pi D_n = \Delta$. En conséquence

$$\Delta' \subset \Delta \subset H.$$

D'ailleurs $H - \Delta' \subset \Sigma R_n$. Car si un point de H n'est pas dans Δ', il n'est pas commun à tous les Γ_m. It est étranger à au moins un Γ_m. Etant dans H, il est dans R_m. Donc l'ensemble $S' = H - \Delta'$ est compris dans ΣR_n. Or $\phi(\Sigma R_n) = 0$. Donc $H = \Delta' + S'$ avec $\phi(S') = 0$.

Δ' étant mesurable-ϕ, il en est de même de H et les deux ensembles H et Δ' ont la même mesure. En vertu des inclusions $\Delta' \subset \Delta \subset H$, Δ est lui aussi mesurable-ϕ et a même mesure que H et que Δ'. Si $H = \Delta + S$, $S \subset S'$ et $\phi(S) = 0$.

15. La première partie de la seconde conclusion du Théorème est établie. Passons à la seconde partie, à savoir la possibilité de choisir les γ_i de façon que $\phi(\Delta) \leqq \phi(\Gamma) < \phi(\Delta) + \epsilon$, le nombre positif ϵ étant préalablement donné.

D'abord la condition $\phi(\Delta) \leqq \phi(\Gamma)$ est vérifiée par toute collection $\Gamma = \Sigma\gamma_i$ telle que $\phi(H - H \cdot \Gamma) = 0$. Car, d'après $H = \Delta + S$ et $\phi(S) = 0$, $\Gamma \cdot H = \Gamma \cdot \Delta + \Gamma \cdot S$, avec $\phi(\Gamma \cdot S) = 0$. Donc $\phi(\Gamma \cdot \Delta) = \phi(\Gamma \cdot H)$. Or $\phi(\Gamma \cdot H) = \phi(H)$, $\phi(H) = \phi(\Delta)$, et évidemment $\phi(\Gamma) \geqq \phi(\Gamma \cdot \Delta)$. Donc $\phi(\Gamma) \geqq \phi(\Delta)$.

Reste à prouver la possibilité que $\phi(\Gamma) < \phi(\Delta) + \epsilon$ pour un choix approprié des γ_i dont la collection est Γ. Evidemment, il suffit pour cela d'établir l'égalité $\lim \phi(\Gamma_n) = \phi(\Delta)$, la collection Γ_n étant celle du paragraphe précédent, composée d'ensembles $\gamma_i{}^n$ de G^n, famille composée des γ vérifiant $\phi(\gamma) < \eta_n$.

D_n est la réunion de ces mêmes γ composant G_n, et Γ_n est donc inclus dans D_n; D_n décroît (plus précisément: ne croît pas) quand n croît, et l'ensemble-limite abstraite de D_n (ensemble commun à tous les D_n, parce que ces ensembles décroissent) est Δ.

Si D_n est mesurable-ϕ, tout au moins à partir d'une certaine valeur de n, $\phi(D_n)$ tend vers $\phi(\Delta)$. D'après $\phi(\Delta) \leqq \phi(\Gamma_n) \leqq \phi(D_n)$, il est évident que $\phi(\Gamma_n)$ tend vers $\phi(\Delta)$.

Le point à établir ne présente de difficulté apparente que si l'on ne suppose pas D_n mesurable-ϕ, ce que *a priori* ne peut être écarté.

Le fait que $\Delta = \lim D_n$ soit mesurable-ϕ, les D_n étant des ensembles décroissants, n'implique nullement ni que D_n soit mesurable-ϕ, ni que $\phi_e(D_n)$ tende vers $\phi(\Delta)$. Il est aisé de former un exemple de cette dernière circonstance, et avec la mesure euclidienne. On sait que l'on peut décomposer l'intervalle $\sigma(0,1)$ des nombres réels en une infinité dénombrable d'ensembles E_n disjoints et dont les longueurs extérieures $l_e(E_n)$ sont toutes égales à 1, leurs longueurs intérieures étant dês lors toutes nulles.[5] Posons $\sigma_n = \sum_{m \geqq n} E_m$, en sorte que $\sigma = \sigma_1$. Dès $n \geq 2$, σ_n, ayant comme chacun des E_m, pour $m \geq n$, la mesure extérieure 1, tandis que son complémentaire $\sigma - \sigma_n$, a lui aussi, comme chacun des E_i pour $i < n$ la même mesure extérieure 1, σ_n est non mesurable. Quand n croît, σ_n décroît, et tout point de σ cesse à partir d'une certaine valeur de n de faire partie de σ_n, puisque ce point appartient à un des E_i et à un seul. Donc l'ensemble commun aux σ_n, soit $\lim . \sigma_n$, est vide. Or $l_e(\sigma_n) = 1$ indépendamment de n; $l_e(\sigma_n)$ ne tend pas vers $l(\lim \sigma_n) = 0$.

---

[5] Voir par exemple Denjoy, *Leçons sur le calcul des coefficients des séries trigonométriques*, 2e partie, pp. 146-147.

Le passage cité appelle toutefois un développement complémentaire pour permettre d'affirmer que les E_n, tous situés sur σ, ont tous pour mesure extérieure l'unité. Rappelons les considérations initiales.

On énumère en une suite unique $\theta_1, \theta_2, \cdots, \theta_n, \cdots$ les nombres rationnels de l'intervalle $\sigma(0,1)$. Soit η leur ensemble. x étant un point quelconque de σ, désignons par $e(x)$ l'ensemble formé de x et des points x' de σ tels que $x' - x$ $(x < x')$ ou $x' - x + 1$ $(x' < x)$ soit dans η; $e(x + \theta_n)$ si $x + \theta_n < 1$, $e(x + \theta_n - 1)$ si $x + \theta_n > 1$ sont identiques à $e(x)$. Soit E_0 un ensemble contenant un point et un seul de chacun des ensembles $e(x)$, et E_n l'ensemble des $x + \theta_n$ $(x + \theta_n < 1)$ ou $x + \theta_n - 1$ $(x + \theta_n > 1)$, x décrivant E_0. Les E_n sont disjoints. Ils ont tous la mesure intérieure zéro, la même mesure extérieure β positive. Seulement *a priori*, β n'est pas nécessairement égal · 1 (β peut être supposé aussi petit que l'on veut, E_0 pouvant être placé sur l'intervalle $0, \beta$).

Il existe sur σ un ensemble G_δ de mesure β contenant E_0 et, si $0 < \lambda < \beta$, il y a dans G_δ un ensemble parfait P, épais en lui-même, de mesure λ. L'ensemble $E_0 \cdot P = I_0$ a pour mesure extérieure λ (sans quoi E_0 n'aurait pas pour mesure extérieure β). Soit $u_1, u_2, \cdots, u_k, \cdots$ la suite des intervalles contigus à P, rangés par longueurs non croissantes. Le nombre $s_k = k u_k + \sum_{p \geqq k+1} u_p$ tend vers 0 si k augmente. Imprimons à P une translation rationnelle θ_n inférieure à u_k. L'ensemble parfait obtenu P_k est tel que $P \cdot P_k$ a une longueur supérieure à $\lambda - s_k$. Soit $k_1, k_2, \cdots, k_m, \cdots$ une suite d'entiers croissants tels que la série $\sum s_{k_m}$ converge et ait une somme inférieure à λ. Si P_m et I_m sont respectivement déduits de P et de I_0 par une translation θ_{n_m} inférieure à u_{k_m}, d'une part $I_m = E_{n_m} \cdot P_m$, d'autre part λ est à la fois la longueur de P_m et la longueur extérieure de I_m. Si Q est l'ensemble majeur épais en lui-même commun aux P_m, la longueur μ de Q est positive et Q est parfait. Les ensembles $J_m = I_m \cdot Q = E_{n_m} \cdot Q$ sont disjoints et ont la mesure extérieure μ. Si a est l'origine de Q et $m(x)$ la longueur de Q entre a et x $(> a)$, la transformation $y = m(x)/\mu$ change Q en le segment $(0,1)$ et les ensembles J_m en des ensembles H_m situés sur σ, disjoints et tous de mesure intérieure zéro, de mesure extérieure 1.

16. Supposons que D_n soit non mesurable-ϕ; $D_n - \Delta$ est non mesurable-ϕ. Cet ensemble peut se décomposer en deux ensembles disjoints (le premier éventuellement vide) X_n et Y_n tels que: 1° X_n est mesurable-ϕ et $\phi(X_n)$ a la plus grande valeur possible, en sorte que $\phi(X_n) = \phi_i(D_n - \Delta)$; 2° Y_n est non mesurable-ϕ, sa mesure intérieure $\phi_i(Y_n)$ étant nulle d'après la valeur de $\phi(X_n)$; D_n est décomposé en trois ensembles disjoints $D_n = \Delta + X_n + Y_n$ et $\phi_e(D_n) = \phi(\Delta) + \phi(X_n) + \phi_e(Y_n)$, $\phi_i(D_n) = \phi(\Delta) + \phi(X_n)$.

On peut supposer que X_n décroît (ne croît pas) quand n croît. Car X_{n+1} inclus dans D_{n+1} l'est *a fortiori* dans D_n. Les points de X_{n+1} étrangers à X_n forment un ensemble mesurable-ϕ dont la mesure-ϕ est nulle. Sans quoi, en ajoutant ces points à X_n, on augmenterait $\phi(X_n)$, sans que X_n cessât d'être inclus dans D_n, ce qui est impossible, puisque $\phi(X_n) = \phi_i(D_n - \Delta)$ est par hypothèse le maximum des mesures-ϕ des ensembles mesurables-ϕ inclus dans $D_n - \Delta$. En retranchant de X_{n+1} les points étrangers à X_n pour les ajouter à Y_{n+1}, on n'altere pas la condition $\phi(X_{n+1}) = \phi_i(D_{n+1} - \Delta)$, et l'on réalise l'inclusion $X_{n+1} \subseteq X_n$.

X_n décroît. Comme $X_n + Y_n$ égal à $D_n - \Delta$ décroissant (non-croissant) a pour limite abstraite (ensemble commun) l'ensemble vide, pareillement $\lim X_n$ est vide, et $\phi(X_n)$ tend vers 0 quand n croît.

Or, Γ_n étant dans D_n, $\Gamma_n = \Gamma_n \cdot D_n = \Gamma_n \cdot \Delta + \Gamma_n \cdot X_n + \Gamma_n \cdot Y_n$; Γ_n est mesurable-ϕ, comme somme des γ_i eux-mêmes mesurables-ϕ par hypothèse. De même, $\Gamma_n \cdot \Delta$ et $\Gamma_n \cdot X_n$ sont mesurables-ϕ puisque Δ et X_n le sont. Donc $\Gamma_n \cdot Y_n$ est mesurable-ϕ, et comme cet ensemble est inclus dans Y_n dont la mesure intérieure-ϕ est nulle, $\phi(\Gamma_n \cdot Y_n) = 0$. En conséquence $\phi(\Gamma_n) = \phi(\Gamma_n \cdot \Delta) + \phi(\Gamma_n \cdot X_n)$.

$\phi(\Gamma_n \cdot X_n)$ tend vers 0 puisqu'il en est ainsi de $\phi(X_n)$; $\phi(\Gamma_n \cdot \Delta) = \phi(\Delta)$, indépendamment de n. Donc $\lim \phi(\Gamma_n) = \phi(\Delta)$.

ϵ étant un nombre positif donné, il est donc possible de choisir η_n assez petit pour que la collection Γ_n ou Γ composée avec des ensembles disjoints γ_i de la famille définie par $\phi(\gamma) < \eta_n$ vérifie $\phi(\Gamma) < \phi(\Delta) + \epsilon$.

La seconde conclusion du Théorème Général est entièrement démontrée. La proposition complète est donc établie.

17. Sur les trois hypothèses conditionnant le théorème général, on peut se demander si la troisième $[\phi_e(\Sigma\gamma) < \infty]$ est indispensable. Un exemple réalisé dans le plan U_2, avec la métrique de l'aire euclidienne des ensembles, va nous montrer un couple de familles P, G d'ensembles ω et γ respectivement, vérifiant les deux premières hypothèses du théorème, l'ensemble H indéfiniment recouvert, au sens des aires, par la famille G des ensembles ω ayant une aire déterminée et finie, tandis que la conclusion du théorème général est mise en défaut. Non seulement H ne peut pas être couvert par une collection dénombrable d'ensembles γ, mais les γ sont tous disjoints de H.

Regardons U_2 comme le plan complexe de la variable $z = x + iy = re^{i\theta}$. Soit C le cercle ouvert $r < 1$; $\epsilon_1, \epsilon_2, \cdots, \epsilon_n, \cdots$ étant une suite de nombres positifs décroissants tendant vers 0, avec $\epsilon_1 < 1/4$, nous formons une série de familles associées, P_n d'ensembles ω et G_n d'ensembles γ, de la façon suivante. $z = re^{i\theta}$ étant un point quelconque de C, nous constituons un ensemble ω de P_n par les deux cercles de même rayon ϵ_n et de centres respectifs z et $z + 2ne^{i\theta} = (r + 2n)e^{i\theta}$. L'ensemble $\gamma = \gamma(\omega)$ correspondant dans G_n sera le second de ces cercles. Les familles P et G sont respectivement ΣP_n et ΣG_n.

Les cercles γ de G_n forment par leur réunion la couronne circulaire I_n $(2n - \epsilon_n < r < 2n + 1 + \epsilon_n)$. La distance de deux de ces couronnes consécutives et celle de I_1 à I_0 $(0 < r < 1 + \epsilon_1)$ surpassent 1/2.

L'ensemble H indéfiniment couvert par la famille G, au sens des aires (et de la dimension, et au sens de toute métrique donnant la mesure zéro à tout point et une mesure positive à tout ensemble ouvert) est le cercle fermé $\bar{C}$ $(r \leqq 1)$. L'ensemble $D = \Sigma\gamma$ est $\sum_{n \geqq 1} I_n$; D est distant de H d'au moins 3/4. Telle est la distance minimum de tout ensemble γ à H.

Or les deux premières hypothèses du théorème général sont vérifiées par les familles P, G. En effet, si un ensemble $\omega' = \omega(\gamma')$ est joint à un γ et si celui-ci est dans G_n, ω' est joint à γ par sa partie $\gamma(\omega')$, uniquement, puisque $\omega' - \gamma(\omega')$ est dans I_0. Donc $\gamma(\omega')$ est dans I_n, comme γ; ω' est dans P_n comme $\omega(\gamma)$. Il suit de là que, les ω' joints à γ, ayant leur aire bornée inférieurement par $2\pi\epsilon_n^2$, ne recouvrent indéfiniment aucun point, au sens des aires (ni d'ailleurs au sens d'aucune métrique donnant une mesure positive à tout ensemble ouvert). Donc $\rho(\gamma)$ est vide quel soit γ. La première hypothèse $\phi[\rho(\gamma)] = 0$ est vérifiée.

[Pour associer aux conclusions précédentes une métrique borélienne $\phi(E)$ plus générale que celle des aires euclidiennes, il conviendrait d'exiger plus précisément que, si c est un cercle de rayon r, $\phi(c)$ vérifie une double inégalité $0 < h(r) < \phi(c) < k(r)$, h et k étant indépendants de c et $k(r)$ tendant vers 0 avec r].

Passons à la seconde. Quel que soit $a > 1$, les ω' vérifiant $\phi(\gamma') < a\phi(\gamma)$, si $\gamma' = \gamma(\omega')$, et joints à γ sont, en raison de cette dernière condition, dans P_n. Ils s'identifient à tous les ω' joints à γ. Pour eux $\phi(\gamma') = \phi(\gamma)$, avec la métrique-$\phi$ des aires. Chacun de ces ω' se compose de deux cercles de même rayon ϵ_n que γ. Leur réunion $\Omega(\gamma)$ est contenue dans un couple de cercles ayant le même rayon $3\epsilon_n$ et respectivement concentriques aux deux cercles composant $\omega(\gamma)$. Donc $\Omega(\gamma)$ a son aire inférieure à 18 fois l'aire de γ. La seconde hypothèse est vérifiée.

Mais la troisième: aire $D < \infty$ ne l'est pas puisque $D = \Sigma I_n$, l'aire de I_n surpassant $4\pi n$. Cette troisième hypothèse joue donc un rôle indispensable dans le théorème général.

18. Un cas particulier important est celui où les ensembles ω se confondent avec les ensembles γ. C'est par l'examen de cette hypothèse que nous terminerons cette étude.

Nous dirons que, dans l'espace U où a été définie une métrique $\phi(E)$ des ensembles, borélienne, non négative, une *famille* G d'ensembles γ est *régulière au sens de la métrique-*ϕ si:

1° *Tout ensemble* γ *a une mesure-*ϕ *déterminée positive et finie;*

2°. *Quel que soit l'ensemble* γ *de* G, *les points étrangers à* γ *et couverts indéfiniment, au sens de la métrique-*ϕ, *par les* γ' *joints à* γ *forment un ensemble* $\rho(\gamma)$ *de mesure-*ϕ *nulle;*

3°. *Il existe deux nombres* a *et* b $(1 < a < b)$ *indépendants de* γ, *tels que la réunion* $\Omega(\gamma)$ *des* γ' *joints à* γ *et vérifiant* $\phi(\gamma') < a\phi(\gamma)$ *a une mesure extérieure-*ϕ *inférieure à* $b\phi(\gamma)$: $\phi_e[\Omega(\gamma)] < b\phi(\gamma)$;

4°. *La réunion des* γ *est un ensemble* D *de mesure extérieure finie*: $\phi_e(D) < \infty$.

Nous désignons par $\Delta(G)$ *l'ensemble indéfiniment couvert, au sens de la métrique-*ϕ, *par la famille* G *des ensembles* γ.

Le théorème général (10), où l'on identifie les ω avec les γ, donne l'énoncé suivant:

THÉORÈME. *L'ensemble* $\Delta(G)$ *relatif à la famille* G *d'ensembles* γ, *régulière au sens de la métrique-*ϕ, *peut être couvert, à un ensemble de mesure-*ϕ *nulle près, par une collection dénombrable* $\Gamma = \Sigma\gamma_i$ *d'ensembles* $\gamma_1, \gamma_2, \cdots$, *disjoints, appartenant à* G, *et* $\Delta(G)$ *est mesurable-*ϕ. En outre, quelque soit le nombre *positif* ϵ préalablement donné, *on peut toujours réaliser l'égalité* $\phi(\Gamma) < \phi(\Delta) + \epsilon$.

Il est à peine besoin de noter que, dans l'espace cartésien U_r à r dimensions et si la métrique $\phi(E)$ est la mesure euclidienne borélienne d'ordre r, on a une famille régulière dans tout groupement, borné dans U_2, d'ensembles γ ayant une mesure déterminée d'ordre r, si cette dernière surpasse une fraction fixe (indépendante de γ) de d^r, d étant le diamètre de γ. Par exemple, les γ peuvent être des sphères, ou des quasi-parallèlipipèdes rectangles d'ordre r, dont les côtés ont des rapports mutuels compris entre deux nombres fixes k et $1/k$ $(0 < k < 1)$. C'est à cette dernière sorte d'ensembles que se rapporte le théorème classique de Vitali. Mais notre définition des familles régulières englobe évidemment des espèces d'ensembles bien plus générales, où la régularité de l'assemblage doit être assurée dans chaque cas particulier par la consistance de la famille considérée.

19. Pour le dernier théorème (18) on peut se poser la même question que pour le Théorème général: Si les ensembles γ de la famille G vérifient les deux premières hypothèses, et si l'on suppose en outre que l'ensemble $\Delta(G)$ indéfiniment couvert par les γ a une mesure extérieure finie, la troisième hypothèse est-elle ou non superflue?

Dans le plan U_2 et avec la mesure euclidienne des aires, soit C_0 le carré ouvert $(0 < x < 1, 0 < y < 1)$. Divisons-le en quatre carrés égaux, et chacun de ceux-ci en neuf autres dont les carrés centraux fermés seront notés C_1^1. Nous avons partagé C_0 en 36 carrés égaux dont 4 ont été retenus. Il en reste 32, se partageant en quatre groupes de huit entourant un même carré C_1^1. Dans chacun de ces 32 carrés complémentaires opérons comme nous l'avons fait sur C_0: division en 36 carrés égaux sur lesquels 4 sont retenus et notés C_1^2. Nous répétons indéfiniment l'opération. Nous avons des collections de 4 carrés C_1^1, $4.\,32$ carrés C_1^2, $4.\,32^2$ carrés $C_1^3, \cdots, 4.\,32^{n-1}$ carrés C_1^n, etc. Les carrés C_1^n sont disjoints et leur réunion $\Sigma C_1^n = \Lambda^1$ couvre C_0 moins un ensemble d'aire nulle. Car l'aire de C_0 diminué des carrés C_1^1 vaut $8/9$, l'aire de C_0 moins les C_1^1 et les C_1^2 vaut $(8/9)^2, \cdots$. L'aire de $C_0 - \Sigma C_1^1 - \Sigma C_1^2 \cdots - \Sigma C_1^n$ vaut $(8/9)^n$ et tend vers 0 quand n croît.

Maintenant, sur chacun des C_1^n $(n = 1, 2, \cdots)$ procédons comme sur C_0. Nous obtenons une famille Λ_2 de carrés fermés disjoints C_2^p qui couvrent chacun des C_1^n à un ensemble d'aire nulle près et C_0 de la même manière. Et nous recommençons avec les C_2^p pour obtenir une famille Λ_3 de carrés C_3^q disjoints, couvrant C_0 à un ensemble d'aire nulle près, etc. La réunion des familles Λ_m est une famille Λ d'ensembles fermés C_m^k couvrant indéfiniment un ensemble Υ qui est C_0 diminué d'un ensemble d'aire nulle (formé des points se projetant sur l'un ou l'autre de deux ensembles égaux, sommes d'ensembles parfaits totalement discontinus privés de leurs points de première espèce et situés respectivement sur Ox et sur Oy).

La question qui se pose est celle-ci. Peut-on adjoindre à chacun des C_m^k

un ensemble g_m^k extérieur à C_0, et par exemple fermé, la réunion $C_m^k + g_m^k$ formant un ensemble γ (ou $\gamma_{m,k}$), de façon que, ces trois conditions étant vérifiées:

1°. l'ensemble Δ indéfiniment couvert, au sens des aires, par la famille G des γ est tout entier dans C_0, donc identique à Υ (l'ensemble indéfiniment couvert par la famille des g_m^k est vide);

2° et 3°. les γ vérifient les deux premières hypothèses,

néanmoins il soit impossible de couvrir $\Delta \equiv \Upsilon$ (moins un ensemble d'aire nulle) avec une infinité dénombrable de γ_i disjoints, parce que, pour aucun choix d'une collection de C_m^k disjoints et couvrant C_0 sauf un ensemble d'aire nulle, les g_m^k correspondants ne sont tous disjoints?

Bien entendu cette impossibilité exige que l'aire de $D = \Sigma\gamma$ (donc celle de Σg_m^k) soit infinie.

Je laisse au lecteur le soin d'élucider ce point.

20. Il est évident que *toute famille partielle d'une famille régulière G est elle-même régulière,* toujours au sens de la métrique-ϕ.

On peut se poser la question suivante. Si $\Delta(G)$ est décomposé en deux ensembles mesurables-ϕ disjoints, H' et H'', est-il toujours possible de trouver dans la famille G deux familles disjointes G' et G'' d'ensembles γ de façon que $\Delta(G') = H'$, à un ensemble de mesure-ϕ nulle près, et sous la même réserve $\Delta(G'') = H''$?

Il n'en est rien, en voici un exemple. Dans le plan U_2 où la métrique est celle des aires boréliennes, soient C_1 et C_2 deux couronnes circulaires ouvertes, homothétiques par rapport à leur centre commun 0 et disjointes, par exemple C_1 de rayon 1 et 2, C_2 de rayons 3 et 6. Prenons pour ensembles γ tous les couples de cercles c_1, c_2, homothétiques entre eux avec le rapport de 1 à 3, par rapport à 0, et respectivement intérieurs à C_1 et à C_2. On voit immédiatement que la famille G des γ vérifie les quatre conditions de régularité. Si les deux ensembles $\gamma = c_1 + c_2$ et $\gamma' = c'_1 + c'_2$ de G sont joints, simultanément c_1 et c'_1 d'une part, c_2 et c'_2 d'autre part sont joints. D'où résultent immédiatement les second et troisième caractères de régularité de la famille G des γ. Le premier et le quatrième caractères sont évidemment réalisés aussi. $\Delta(G)$ est identique à $C_1 + C_2$.

Toute famille partielle de G couvre deux ensembles distincts, l'un situé dans C_1, l'autre dans C_2, le second homothétique du premier par rapport à 0, avec le rapport 3. Donc, quel que soit l'ensemble H_1 dans C_1, si son homothétique par rapport à 0 avec le rapport 3 est l'ensemble H_2 situé dans C_2, toute famille partielle de G couvrant indéfiniment H_1 couvre en même temps H_2. Et réciproquement. En particulier, si $\Delta(G) = C_1 + C_2$ est décomposé en $H' = C_1$ et $H'' = C_2$, toute famille G' partielle de G couvrant indéfiniment H' et toute famille G'' partielle de G couvrant indéfiniment H'', toujours au sens des aires, et à un ensemble d'aire nulle près, donnent lieu aux identités: $\Delta(G') = \Delta(G'') = \Delta(G) = C_1 + C_2$, à un ensemble d'aire nulle près.

Cet exemple, qui pourrait être transformé dans le détail (les γ de G composés de couples de cercles c_1, c_2 homothétiques avec le rapport 3 ne couvrant que deux régions homothétiques de C_1 et de C_2) fait saisir pourquoi, contrairement à ce que l'on pourrait supposer, les conclusions des théorèmes qui suivent n'admettent pas en général la substitution aux ensembles $\Delta(G)$ de n'importe laquelle de leurs parties mesurables.

21. Ce sont les familles régulières G telles que les précédentes qui donnent de l'intérêt à la proposition suivante:

THÉORÈME. *Au sens de la métrique-ϕ, si l'ensemble H de mesure extérieure-ϕ positive est inclus dans l'ensemble $\Delta = \Delta(G)$ indéfiniment couvert par la famille régulière G d'ensembles γ, et si H vérifie, pour tout ensemble γ de G l'inégalité $\phi_e(H \cdot \gamma) < \theta\phi(\gamma)$, θ étant inférieur à 1 et indépendant de γ, l'inégalité $\phi(\Delta) < \theta^{-1}\phi_e(H)$ est impossible.*

Soit en effet, $\Gamma = \Sigma\gamma_i$ l'ensemble-somme d'une collection dénombrable d'ensembles $\gamma_1, \gamma_2, \cdots$ de G, disjoints et couvrant Δ à un ensemble de mesure-ϕ nulle près. En outre, ϵ désignant (et nous ne le répèterons plus dans les pages suivantes) un nombre positif indépendant, on peut réaliser la condition $\phi(\Gamma) < \phi(\Delta) + \epsilon$ (15-16).

De l'égalité $\Delta = \Delta \cdot \Gamma + \Theta$ avec $\phi(\Theta) = 0$, et de la relation d'inclusion $H \subset \Delta$, il résulte: $H = H \cdot \Delta = H \cdot \Gamma + H \cdot \Theta$ et $\phi(H \cdot \Theta) = 0$. Donc $\phi_e(H) = \phi_e(H \cdot \Gamma)$. Les ensembles γ_i étant disjoints (6), $\phi_e(H \cdot \Gamma) = \Sigma\phi_e(H \cdot \gamma_i)$. Donc $\phi_e(H) < \theta\Sigma\phi(\gamma_i) = \theta\phi(\Gamma) < \theta[\phi(\Delta) + \epsilon]$; ϵ étant positif quelconque, $\phi_e(H) \leq \theta\phi(\Delta)$. L'inégalité $\phi(\Delta) < \theta^{-1}\phi_e(H)$ est impossible.

22. Les espaces généraux U, avec les métriques quelconques, donnent lieu à un théorème coincidant pour les espaces cartésiens et la métrique euclidienne borélienne, les familles G étant composées de sphères, avec le théorème sur l'*épaisseur* des ensembles mesurables.[6]

---

[6] J'emploie le terme d'*épaisseur* où trop d'analystes usent de celui de *densité*. Je

Dans U soit G une famille d'ensembles γ. Au sens de la métrique-ϕ, supposons G régulière et couvrant indéfiniment l'ensemble $\Delta = \Delta(G)$. Soit E un ensemble mesurable-ϕ inclus dans Δ, M un point de Δ et $\eta_i(E, M)$, $\eta_s(E, M)$ respectivement *la plus petite* et *la plus grande limites du rapport* $\phi(E \cdot \gamma)/\phi(\gamma)$ *quand l'ensemble* γ *de* G *contient* M et *que sa mesure* $\phi(\gamma)$ *tend vers* 0; $\eta_i(E, M)$, $\eta_s(E, M)$ seront appelées respectivement *l'épaisseur inférieure, l'épaisseur supérieure de l'ensemble* E *au point* M, *relativement à la famille* G. Si ces deux nombres sont égaux, leur valeur commune sera appelée *l'épaisseur de* E *au point* M *relativement à la famille* G.

THÉORÈME. *Au sens de la métrique-*ϕ*, si la famille* G *d'ensembles* γ *est régulière et si* $\Delta = \Delta(G)$ *est l'ensemble indéfiniment couvert par elle, l'épaisseur de* Δ *relativement à la famille* G *est* 1 *en tout point de* Δ, *sauf éventuellement sur un ensemble de mesure-*ϕ *nulle.*

L'hypothèse contraire est que l'ensemble des points M vérifiant $\eta_i(\Delta, M) < 1$ a une mesure extérieure-ϕ positive. Considérons l'ensemble des M vérifiant $\eta_i(\Delta, M) < 1 - 1/n$. Si, quel que fût n, cet ensemble avait la mesure-ϕ nulle, l'ensemble $\eta_i(\Delta, M) < 1$ aurait aussi la mesure 0. Il y a donc un nombre $\theta < 1$ tel que l'ensemble H des points M vérifiant $\eta_i(\Delta, M) < \theta$ a sa mesure extérieure $\phi_e(H)$ positive.

Soit G' la famille des ensembles γ appartenant à G et vérifiant $\phi(\Delta \cdot \gamma) < \theta\phi(\gamma)$. L'ensemble $\Delta' = \Delta(G')$, indéfiniment couvert par la famille G' contient H. On peut d'ailleurs observer qu'en tout point de $\Delta' - H$, on trouve $\eta_i(\Delta, M) = \theta$; Δ' est mesurable-ϕ (14). Donc $\phi(\Delta') \geq \phi_e(H) > 0$. Couvrons Δ', à un ensemble de mesure-ϕ nulle près, avec une collection dénombrable d'ensembles $\gamma'_1, \gamma'_2, \cdots$ de G', disjoints, et soit $\Gamma' = \Sigma\gamma'_i$. On peut supposer $\phi(\Gamma') < \phi(\Delta') + \epsilon$. (15-16). D'après $\Delta' = \Delta' \cdot \Gamma' + \Theta'$, avec $\phi(\Theta') = 0$, $\phi(\Delta') = \phi(\Delta' \cdot \Gamma')$. L'inclusion $\Delta' \subset \Delta$ donne $\Delta' \cdot \Gamma' \subset \Delta \cdot \Gamma'$ et $\phi(\Delta') \leq \phi(\Delta \cdot \Gamma')$. Mais, les γ'_i étant disjoints

$$\phi(\Delta \cdot \Gamma') = \Sigma\phi(\Delta \cdot \gamma'_i) < \theta\Sigma\phi(\gamma'_i) = \theta\phi(\Gamma') < \theta[\phi(\Delta') + \epsilon].$$

En conséquence $\phi(\Delta') < \theta[\phi(\Delta') + \epsilon]$, quel que soit ϵ positif. Donc

$$\phi(\Delta') \leq \theta\phi(\Delta'),$$

---

réserve celui-ci à la notion topologique, fondamentale dans la théorie des fonctions de variables réelles.

Voir mes *Leçons sur le Calcul des coefficients des séries trigonométriques* (note liminaire p. XI, pp. 105-106 en note; p. 223 en note). Au sujet de la densité, voir également: p. 136 en note, 331 en note. Et consulter la Critique du Langage, pp. 689-690.

ce qui est absurde si, avec $\theta < 1$, en même temps $\phi(\Delta') > 0$. Or il en est ainsi au cas où $\phi_e(H) > 0$.

23. APPLICATIONS AUX FONCTIONS D'ENSEMBLES. Dans l'espace U pourvu de la métrique-ϕ, soit $\psi(E)$ une fonction d'ensemble complètement additive, et même au sens précis suivant: Si les ensembles $E_1, E_2, \cdots$ sont disjoints, si tous les $\psi(E_n)$ sont déterminés et si la série $\Sigma \mid \psi(E_n) \mid$ converge, $\psi(E)$ existe et vaut $\Sigma\psi(E_n)$. En outre, on suppose la soustractivité de $\psi(E)$: Si, avec $E \supset E'$, $\psi(E)$ et $\psi(E')$ sont définis, $\psi(E - E')$ existe et dès lors vaut $\psi(E) - \psi(E')$.

Enfin (6), dès que $\psi(E)$ et $\psi(E')$ sont définis, $\psi(E \cdot E')$ l'est aussi.

Dans U soit G une famille d'ensembles γ. Au sens de la métrique-ϕ, nous supposons G régulière et couvrant indéfiniment l'ensemble $\Delta = \Delta(G)$; Δ est mesurable-ϕ (14). Nous dirons que la fonction $\psi(E)$ est *définie dans le champ* $\Delta(G)$ si $\psi(E)$ est déterminé et fini pour tout ensemble H mesurable-ϕ inclus dans Δ (y compris Δ). Nous ne rappellerons plus désormais les hypothèses générales ni le sens des notations adoptées.

Soit M un point quelconque de Δ. Nous appelerons *nombres dérivés supérieur, inférieur, moyen, relatifs à la famille G et propres à la fonction d'ensemble ψ au point M,* respectivement la plus grande, la plus petite, une quelconque des limites du rapport $\psi(\Delta \cdot \gamma)/\phi(\gamma)$ si l'ensemble γ de G contient M et si $\phi(\gamma)$ tend vers 0.

Nous noterons ces nombres $\Lambda_s(\psi, M, G)$, $\Lambda_i(\psi, M, G)$, $\Lambda_m(\psi, M, G)$.

Nous dirons que $\psi(E)$ possède *au point M une dérivée relative à la famille régulière G* si $\Lambda_s(\psi, M, G) = \Lambda_i(\psi, M, G)$. Cette dérivée sera notée $\Lambda(\psi, M, G) = \Lambda_i = \Lambda_s$.

Je dis que *les nombres dérivée Λ_i et Λ_s sont mesurables-ϕ.* Une fonction de point $f(M)$ est dite *mesurable-ϕ* si, H désignant son champ de définition dans U, l'ensemble des M où $A < f(M) < B$ (M inclus dans H) est mesurable-ϕ, quels que soient A et B ($A < B$).

Soit A un nombre réel quelconque et I_A l'ensemble où $\Lambda_i(\psi, M, G) < A$. Si nous formons la famille $G_i(A)$ des ensembles γ de G pour lesquels $\psi(\Delta \cdot \gamma) < A\phi(\gamma)$, l'ensemble $\Delta[G_i(A)]$ indéfiniment couvert par cette famille renferme I_A et les points de $\Delta[G_i(A)]$ étrangers à A vérifient $\Lambda_i(\psi, M, G) = A$.

Or cet ensemble $\Delta[G_i(A)]$ est mesurable-ϕ, puisque la famille $G_i(A)$, partie de la famille régulière G, est elle-même régulière au sens de la métrique-ϕ. Donc, quel que soit A, il y a un ensemble I'_A, savoir $\Delta[G_i(A)]$, mesurable-ϕ et qui d'une part contient l'ensemble $\Lambda_i(\psi, M, G) < A$, d'autre part est compris dans l'ensemble $\Lambda_i(\psi, M, G) \leq A$. On en conclut, parle raisonnement de Lebesgue relatif aux fonctions mesurables d'une variable réelle, que la fonction $\Lambda_i(\psi, M, G)$ est mesurable-ϕ.

Et de même la fonction $\Lambda_s(\psi, M, G)$ est mesurable-ϕ.

L'ensemble J des points où $\psi(E)$ a une dérivée Λ relative à la famille G est mesurable. Car d'une part $\Lambda_s - \Lambda_i$ est mesurable, et d'autre part J est l'ensemble où $\Lambda_s - \Lambda_i = 0$. Sur l'ensemble J mesurable-ϕ la dérivée Λ est identique à l'une et à l'autre des fonctions Λ_s et Λ_i, toutes deux mesurables. Donc Λ, définie uniquement sur J, est mesurable-ϕ.

24. THÉORÈME. *Si la fonction d'ensemble $\psi(E)$ est définie et bornée dans le champ $\Delta = \Delta(G)$, l'ensemble I des points de $\Delta(G)$ où l'un des nombres dérivés extrêmes de $\psi(E)$ relativement à la famille régulière G d'ensembles γ est infini, I a la mesure-ϕ nulle.*

Par hypothèse, il existe un nombre positif k tel que $-k < \psi(E) < k$, quel que soit E mesurable-ϕ et inclus dans $\Delta(G)$.

Soit H l'ensemble $\Lambda_s(\psi, M, G) = +\infty$; H est mesurable-ϕ. Supposons $\phi(H) > 0$ et montrons-en l'impossibilité. Soit $G_s(A)$ la famille des γ vérifiant $\psi(\Delta \cdot \gamma) > A\phi(\gamma)$. Elle est régulière comme l'est G. L'ensemble $\Delta[G_s(A)]$, ou par abréviation $\Delta_s(A)$, contient H quel que soit A. Il existe une réunion dénombrable Γ_A d'ensembles γ_i de $G_s(A)$ disjoints, tels que

$$\Delta_s(A) - \Delta_s(A) \cdot \Gamma_A = \Theta_A \text{ a la mesure-}\phi \text{ nulle.}$$

Si $L = \Delta - \Delta \cdot \Gamma_A$, $\psi(\Delta) \geq \psi(\Delta \cdot \Gamma_A) - k$; $\psi(\Delta \cdot \gamma_i) > A\phi(\gamma_i)$ est positif quel que soit i. Les γ_i étant disjoints $\sum\limits_{i<n} \psi(\Delta \cdot \gamma_i)$ est inférieur à k quel que soit n. Donc la série $\Sigma\psi(\Delta \cdot \gamma_i)$ converge et vaut $\psi(\Delta \cdot \Gamma_A)$. En conséquence $\psi(\Delta \cdot \Gamma_A) > A\Sigma\phi(\gamma_i) = A\phi(\Gamma_A)$ et $\psi(\Delta) > A\phi(\Gamma_A) - k$.

Or, Γ_A contient $\Delta_s(A)$ sauf un ensemble de mesure-ϕ nulle. Donc $\phi(\Gamma_A) \geq \phi[\Delta_s(A)]$; $\Delta_s(A)$ contient H. Donc $\phi(\Gamma_A) \geq \phi(H)$ et $\psi(\Delta) > A\phi(H) - k$; $\phi(H)$ étant positif et A un nombre positif quelconque, nous aboutissons à une impossibilité, puisque $\psi(\Delta) < k$.

On démontre de même que l'ensemble $\Lambda_i(\psi, M, G) = -\infty$ a la mesure-ϕ nulle.

25. THÉORÈME. *Si la fonction $\psi(E)$ est définie dans le champ $\Delta(G)$ et si l'égalité $\phi(E) = 0$ entraîne $\psi(E) = 0$, il n'est pas possible qu'il existe deux nombres réels distincts A et B $(A < B)$ tels qu'en tout point M de $\Delta = \Delta(G)$, $\Lambda_i(\psi, M, G) < A < B < \Lambda_s(\psi, M, G)$.*

On considère les deux familles $G_i(A)$, $G_s(B)$ formées des ensembles γ de G vérifiant respectivement $\psi(\Delta\cdot\gamma) < A\phi(\gamma)$ et $\psi(\Delta\cdot\gamma) > B\phi(\gamma)$. Par hypothèse $\Delta[G_i(A)] = \Delta[G_s(B)] = \Delta(G) = \Delta$.

En recouvrant successivement Δ, sauf un ensemble de mesure-ϕ, nulle (et sur lequel par hypothèse $\psi = 0$) avec deux collections dénombrables d'ensembles γ appartenant respectivement à $G_i(A)$ et à $G_s(B)$, on met en évidence les deux inégalités contradictoires $\psi(\Delta) \leq A\phi(\Delta)$ et $\psi(\Delta) \geq B\phi(\Delta)$.

Nous dirons que la fonction d'ensemble $\psi(E)$, définie et bornée dans le champ $\Delta(G)$ y est *métriquement continue,* si l'ensemble E variant indifféremment dans $\Delta(G)$, la condition $\lim\phi(E) = 0$ entraîne $\lim\psi(E) = 0$.

26. Nous dirons qu'au sens de la métrique-ϕ, la famille régulière G d'ensembles γ est *parfaitement régulière* si, quel que soit l'ensemble H mesurable-ϕ inclus dans $\Delta(G)$, la famille G' des ensembles γ de G joints à H a pour ensemble de recouvrement indéfini $\Delta(G')$ l'ensemble H accru au plus d'un ensemble de mesure-ϕ nulle.

THÉORÈME. *Si, la famille G d'ensembles γ étant parfaitement régulière, la fonction d'ensemble $\psi(E)$ est définie, bornée et métriquement continue dans le champ $\Delta = \Delta(G)$, $\psi(E)$ possède une dérivée finie par rapport à la famille G, sur une plénitude-ϕ de Δ.*

Sinon l'ensemble $\Lambda_s(\psi, G) - \Lambda_i(\psi, G) > 0$ aurait une mesure-ϕ positive. On trouverait alors deux nombres A et B, avec $A < B$, (par exemple $A = p/q$, $B = (p+1)/q$, q entier positif étant assez grand et p ayant une certaine valeur entière) de façon que l'ensemble H où simultanément $\Lambda_i < A$, $\Lambda_s > B$, eût une mesure-ϕ positive. Adoptons cette hypothèse. Nous allons prouver $\psi(H) < A\phi(H)$ et pareillement $\psi(H) > B\phi(H)$. La contradiction sera établie.

Soit G' la famille des ensembles γ de G joints à H et $\Delta' = \Delta(G')$ l'ensemble indéfiniment couvert par la famille G' (au sens le la métrique-ϕ). La famille G étant parfaitement régulière, $\Delta' = H + Z$, avec $\phi(Z) = 0$. La fonction $\psi(E)$ s'annulant sous la condition suffisante $\phi(E) = 0$, on trouve, avec $\phi(H) = \phi(\Delta')$, $\psi(H) = \psi(\Delta')$.

Dans la famille G' retenons uniquement des γ (ils sont joints à H) où $\psi(\Delta\cdot\gamma) < A\phi(\gamma)$. Ces γ forment une famille $G'_i(A)$ qui couvre indéfiniment un ensemble $\Delta_A{}^1 = \Delta[G'_i(A)]$ inclus dans Δ' et contenant H. Donc l'ensemble $Z_1 = \Delta_A{}^1 - H$ vérifie $Z_1 \subset Z$ et $\phi(Z) = 0$. Par suite $\phi(H) = \phi(\Delta') = \phi(\Delta_A{}^1)$ et $\psi(H) = \psi(\Delta') = \psi(\Delta_A{}^1)$.

Couvrons $\Delta_A{}^1$, à un ensemble de mesure nulle près, avec une collection

Γ^1 d'ensembles γ_i^1 disjoints appartenant à $G'_i(A)$. Nous pouvons supposer $\phi(\Gamma^1) < \phi(\Delta_A{}^1) + \epsilon = \phi(\Delta') + \epsilon$, avec $\phi[\Gamma_1 \cdot \Delta_A{}^1] = \phi(\Delta_1{}^A) = \phi(\Delta')$, et $\psi(\Gamma^1 \cdot \Delta_A{}^1) = \psi(\Delta_A{}^1) = \psi(\Delta')$. D'après $\psi(\Delta \cdot \gamma_i^1) < A\phi(\gamma_i^1)$, et $\psi(\Delta \cdot \Gamma^1)$ étant fini, la série $\Sigma\psi(\Delta \cdot \gamma_i^1)$ converge et a pour somme $\psi(\Delta \cdot \Gamma^1)$. Donc $\psi(\Delta \cdot \Gamma^1) = \Sigma\psi(\Delta \cdot \gamma_i^1) < A\Sigma\phi(\gamma_i^1) = A\phi(\Gamma^1) < A[\phi(\Delta') + \epsilon]$.

Soit $X^1 = \Gamma^1 - \Gamma^1 \cdot \Delta_A{}^1$; $\Delta_A{}^1$ et X^1 sont disjoints. Donc

$$\phi(X^1) = \phi(\Gamma^1) - \phi(\Delta') < \epsilon.$$

D'après $\Delta_A{}^1 \subset \Delta' \subset \Delta$ et

$$\Gamma_1 = \Gamma_1 \cdot \Delta_A{}^1 + X^1, \quad \Delta \cdot \Gamma^1 = \Delta_A{}^1 \cdot \Gamma^1 + X^1 \cdot \Delta,$$

les deux ensembles $\Delta_A{}^1$ et X^1 étant disjoints:

$$\psi(\Delta \cdot \Gamma^1) = \psi(\Delta_A{}^1 \cdot \Gamma^1) + \psi(X^1 \cdot \Delta),$$

et puisque

$$\psi(\Gamma^1 \cdot \Delta_A{}^1) = \psi(\Delta'), \quad \psi(\Delta \cdot \Gamma^1) = \psi(\Delta') + \psi(X^1 \cdot \Delta) < A[\phi(\Delta') + \epsilon].$$

En vertu de l'hypothèse que $\psi(E)$ tend uniformément vers 0 quand $\phi(E)$ tend vers 0, variant dans Δ, $\psi(X^1 \cdot \Delta)$ tend vers 0 avec ϵ, puisqu'il en est ainsi de $\phi(X^1 \cdot \Delta) \leqq \phi(X^1) < \epsilon$. A la limite:

$$\psi(\Delta') \leq A\phi(\Delta').$$

Le raisonnement semblable appliqué à la famille $G'_s(B)$ des ensembles γ de G joints à H (donc inclus dans G') et vérifiant $\psi(\Delta \cdot \gamma) > B\phi(\gamma)$ nous conduit à la conclusion $\psi(\Delta') \geq B\phi(\Delta')$, ce qui est contradictoire avec le précédent résultat si $\phi(\Delta') > 0$. Or $\phi(\Delta') = \phi(H)$. Donc $\phi(H) > 0$ est impossible. La dérivée $\Lambda(\psi, G)$ existe sur une plénitude-ϕ de $\Delta = \Delta(G)$, ensemble indéfiniment couvert, au sens de la métrique-ϕ, par la famille parfaitement régulière G.

Pour l'espace cartésien à r dimensions U_r et la mesure euclidienne borélienne d'ordre r, les familles citées plus haut (18) d'ensembles γ surpassant en mesure une fraction fixe de la sphère minimum les contenant, sont parfaitement régulières.

On voit que l'extension donnée au théorème de Vitali, en rapportant toutes les hypothèse à la seule notion d'une fonction métrique borélienne non négative des ensembles, permet d'énoncer, pour tous les espaces où une telle mesure est définissable, et sans aucun postulat de distance entre points ni de topologie, les analogues de certaines propositions fondamentales relatives à la mesure euclidienne dans les espaces cartésiens ou à la dérivation des fonctions d'une variable réelle.

Observations complémentaires.

La correction des épreuves de cet article me donne l'occasion d'ajouter certaines rectifications et quelque développements à l'étude précédente. Le tout a été d'ailleurs signalé dans une note aux Comptes Rendus de l'Académie des Sciences de Paris (13 novembre 1950, t. 231, p. 1013-1015).

Il s'agit des familles régulières G d'ensembles γ. Nous modifions la définition donnée ci-dessus (26) des familles parfaitement régulières et qui n'a pas d'application.

27. Au sujet du caractère de régularité des familles G d'ensembles γ notons un premier point de détail concernant la quatrième condition. Nous exigeons que la mesure *extérieure-ϕ* de l'ensemble $D = \Sigma\gamma$ soit *finie*, mais non pas que D soit mesurable-ϕ.

Etant donné la famille G d'ensembles γ et si $D = D(G) = \Sigma\gamma$ est non mesurable-ϕ, on peut toujours, en retranchant à certains, sinon à chacun, des γ un ensemble de mesure-ϕ nulle, ramener D à être mesurable-ϕ.

En effet, soit $D = D' + Z$, D' étant mesurable-ϕ et vérifiant $\phi(D') = \phi_i(D)$. Donc $\phi_i(Z) = 0$, et $\phi_e(Z) = \phi_e(D - D') > 0$. ($D'$ et Z sont disjoints).

Les propositions suivantes sont évidentes:

L'ensemble obtenu en retranchant d'un noyau un ensemble de mesure-ϕ nulle est encore un noyau. L'ensemble obtenu en ajoutant à une enveloppe un ensemble de mesure-ϕ nulle est encore une enveloppe.

Tout ensemble inclus dans $\Delta(G)$ et de mesure-ϕ nulle est une enveloppe (comme l'est l'ensemble vide). Toute plénitude-ϕ de $\Delta(G)$ est un ensemble-noyau (comme l'est Δ).

γ étant inclus dans D, $\gamma = \gamma \cdot D = \gamma \cdot D' + \gamma \cdot Z$. Selon un raisonnement déjà rencontré (16), nous voyons que, γ et D', donc aussi $\gamma \cdot D'$, étant mesurables-ϕ, $\gamma \cdot Z$ l'est aussi, et $\phi(\gamma \cdot Z) \leq \phi_i(Z) = 0$.

Soit $\gamma' = \gamma - \gamma \cdot Z$; γ' inclus dans γ est mesurable-ϕ et $\phi(\gamma') = \phi(\gamma)$. A toute collection dénombrable $\Gamma = \Sigma\gamma_i$ d'ensembles γ_i disjoints correspond la collection $\Gamma' = \Sigma\gamma'_i$, incluse dans Γ avec $\phi(\Gamma - \Gamma') = 0$. Les deux familles Γ et Γ' couvrent le même ensemble, à un ensemble de mesure-ϕ nulle près. Or $D' = \Sigma\gamma'$ est mesurable ϕ, avec $\phi(D') = \phi_i(D)$.

Néanmoins il est indispensable de conserver à la quatrième condition toute sa généralité. Car, avec une réunion $D = D(G)$ mesurables-ϕ, on peut avoir des familles partielles G' de G, par exemple formées des γ de G joints à un ensemble H inclus dans $\Delta(G)$ et mesurable-ϕ, de façon que la réunion $D(G')$ des γ de G' soit non-mesurable-ϕ.

28. Donnons de ce cas un exemple emprunté à l'espace linéaire U_1 formé par l'axe Ox, avec la métrique borélienne des longueurs. Soient e_0 un ensemble non mesurable situé sur l'intervalle σ_0 $(0, 1)$ (15 en note), e_1 sur $\sigma_1(1, 2)$, e_2 sur $\sigma_2(2, 3)$ les ensembles respectifs $x+1$, $x+2$, si x décrit e_0. Les segments $\alpha\beta$ inclus dans σ_0, définis par les couples de nombres (α, β), peuvent être mis en correspondance par une liaison (A) avec les points de σ_0 chacun à chacun.

Considérons une famille G d'ensembles γ se répartissant comme il suit en trois familles partielles G_1, G_0, G'_0:

1° G_1 est formé des segments s_1 inclus dans σ_1;

2° G_0 est composé des segments s_0 inclus dans σ_0 et correspondant d'après (A) à un point x de $\sigma_0 - e_0$;

3° les s'_0 étant les segments inclus dans σ_0 et correspondant par (A) aux points x de e_0, les ensembles γ de G'_0 sont les segments s'_0 respectivement accrus des deux points $x+1$ et $x+2$.

La famille G est régulière. L'ensemble $D = D(G) = \Sigma\gamma$ est $\sigma_0 + \sigma_1 + e_2$, soit l'intervalle $(0, 2)$ diminué du point 1 et accru de l'ensemble e_2 sur σ_2; D est non mesurable. Avec les notations utilisées plus haut, D se décompose en $D' = \sigma_0 + \sigma_1$ et $Z = e_2$; γ' est $\gamma - \gamma \cdot e_2$; $\gamma \cdot e_2$ n'existe pas si γ est dans $G_1 + G_0$ et alors $\gamma' = \gamma$. Si γ est dans G'_0, $\gamma \cdot e_2$ est le point $x+2$, s'_0 correspondant au point x de e_0; γ' se déduit donc, dans ce cas, de γ par extraction d'un seul point; γ' est la réunion de s'_0 et du point $x+1$ de e_1. Soient G''_0 la famille des $\gamma' = s'_0 + [x+1]$ et $G' = G_1 + G_0 + G''_0$; $D(G')$ est $\sigma_0 + \sigma_1$; $D(G')$ est mesurable. Et $\Delta(G') = \Delta(G) = \sigma_0 + \sigma_1$.

Ainsi, il nous a suffi de retrancher un point à certains des ensembles γ composant G pour obtenir une famille G' d'ensembles γ' dont la réunion $D(G')$ est mesurable, tandis que la réunion $D(G)$ des γ formant G ne l'était pas.

Mais il est facile de trouver dans $\Delta(G')$ un ensemble H de longueur déterminée et tel que si $G(H)$ est la famille des γ' de G' joints à H, la réunion $D[G(H)]$ de ces γ' soit un ensemble non mesurable. Il nous suffit de prendre pour H l'intervalle σ_0. Alors $G(H) = G_0 + G''_0$; $D[G(H)] = \sigma_0 + e_1$ est non mesurable.

En conséquence, pour que toute famille partielle d'une famille régulière soit elle-même régulière, il faut garder au quatrième caractère toute sa généralité.

29. La définition des familles parfaitement régulières donnée dans le texte de l'article (26) est trop restrictive pour être utile.

Prenons un exemple. Dans l'espace cartésien U_r à r dimensions, et avec la mesure euclidienne borélienne d'ordre r, soit G une famille d'ensembles γ consistant en intervalles réguliers ($a_i < x_i < b_i + l$; $i = 1, 2, \cdots, r$; l indépendant de i pour chaque intervalle γ) contenus dans une sphère donnée $D = D(G) = \Delta(G)$. Soit H un ensemble mesurable quelconque situé dans D. Soit $G(H)$ l'ensemble des γ de G joints à H. Si H est fermé, $\Delta[G(H)]$, ensemble des points indéfiniment couverts par les γ de $G(H)$, est identique à H. Mais si H n'est pas fermé, $G(H)$ et $G(\bar{H})$ sont identiques. Si donc la mesure de H est inférieure à celle de son ensemble de fermeture $\bar{H}$, $\Delta[G(H)] - H$ a une mesure positive. Par exemple si H est un ensemble dénombrable partout dense dans G, $\Delta[G(H)] = D$, avec mes $H = 0$.

Avec la définition du texte (26), ne seraient parfaitement régulières que les familles G pour lesquelles $\Delta(G)$ aurait la mesure-ϕ nulle. Aussi renonçons-nous à cette première définition et, lui substituant une nouvelle, nous pourrons conserver le théorème du n° 26, dont la démonstration restera entièrement utilisable. Enfin, tous les ensembles mesurables-ϕ contenus dans $\Delta(G)$ vérifieront le théorème fondamental sur l'épaisseur (22).

30. Ne perdons pas de vue que la fonction métrique $\phi(E)$ est de nature très générale. Ce pourra être la mesure euclidienne d'ordre p de Carathéodory dans l'espace cartésien U_r à r dimensions, r surpassant p.

Par exemple, dans l'espace U_3, prenons pour $\phi(E)$ l'aire de l'ensemble E. Une droite a une aire nulle. L'axe des x peut être enfermé sur le plan des xy dans l'ensemble, ouvert dans ce plan: $|y| < \alpha/(1+x^2)$; α aussi petit qu'on le veut, est proportionnel à l'aire de l'ensemble. Dans U_3 on déformera celui-ci quasi-hélicoïdalement, par une sorte d'enroulement autour de l'axe des x demeurant invariable. Nous obtenons un ensemble $\bar{\omega}$ relatif à Ox. Dans U_3 les droites parallèles à l'un des trois axes de coordonnées et dont les deux coordonnées constantes sont rationnelles ont une aire totale nulle. Ces droites sont en infinité dénombrable. Si on les énumère, on peut inclure la n^e dans un ensemble $\bar{\omega}_n$ du type décrit ci-dessus pour l'axe des x, et d'aire inférieure à ϵ_n, la série $\Sigma\epsilon_n$ étant convergente. Tel est un ensemble γ constituant de G. On envisagera une infinité non dénombrable de tels ensembles. Pour la régularité de la famille G, la première condition, la mesurabilité de γ [avec $\phi(\gamma) > 0$], est vérifiée. Il faudra s'assurer des trois autres caractères. Mais ces perspectives ouvertes sur le cas le plus simple des espaces cartésiens et des mesures euclidiennes doivent rester présentes à l'esprit si l'on veut saisir toute la portée des notions et des énoncés rencontrés ci-après.

31. Quelle que soit la famille régulière (18) G formée d'ensembles γ, nous distinguons dans $\Delta(G)$ deux classes particulières (non nécessairement disjointes) d'ensembles mesurables-ϕ, à savoir les *ensembles-noyaux* et les *ensembles-enveloppes.*

Selon une notation déjà utilisée, si l'ensemble H est inclus dans $\Delta(G)$, nous désignons par $G(H)$ l'ensemble des γ de G joints à H. Au lieu de $\Delta[G(H)]$ nous écrirons $\delta(H)$. C'est donc l'ensemble indéfiniment couvert au sens de la métrique-ϕ par les γ de G joints à H. Evidement, H étant dans $\Delta(G)$, $\delta(H)$ qui est mesurable-ϕ (14), contient H; $\delta(H) \supset H$. Nous designerons par $\sigma(H)$ l'ensemble $\delta(H) - H$. Donce $\delta(H) = H + \sigma(H)$, H et $\sigma(H)$ étant disjoints.

Nous dirons que, relativement à la métrique-ϕ et à la famille régulière G d'ensembles γ, un ensemble F mesurable-ϕ inclus dans $\Delta(G)$, est un *noyau* quand $\sigma(F) = \delta(F) - F$ a la mesure-ϕ nulle.

Quel que soit l'ensemble γ de G, l'ensemble $\gamma \cdot \Delta$ est un noyau. En effet, soit $H = \gamma \cdot \Delta$. Je dis que $\sigma(H) \subset \rho(\gamma)$, $\rho(\gamma)$ étant l'ensemble défini au n° 18.

Soit $d(\gamma)$ l'ensemble indéfiniment couvert (toujours sous-entendu: au sens de la métrique-ϕ) par les γ' de G' joints à γ; $d(\gamma)$ se compose de deux parties disjointes, la première, $d(\gamma) \cdot \gamma$, située dans γ, la seconde $\rho(\gamma)$ étrangère à γ; $d(\gamma)$ étant dans Δ (comme tout ensemble indéfiniment couvert par des ensembles de G), la première partie de $d(\gamma)$ est commune à γ et à Δ. Elle est dans $\gamma \cdot \Delta$. D'ailleurs tout point M de $\gamma \cdot \Delta$ est dans $d(\gamma)$, puisque, étant dans Δ, il est indéfiniment couvert par des ensembles de G, et que ces ensembles contenant M sont joints à γ. Donc $d(\gamma) \cdot \gamma = \gamma \cdot \Delta = H$ et $d(\gamma) = H + \rho(\gamma)$, H et $\rho(\gamma)$ étant disjoints.

Considérons les γ'' de G composant $G(H)$. Etant joints à H qui est dans γ, ils sont joints à γ. Donc l'ensemble $\delta(H)$ indéfiniment couvert par les γ'' est inclus dans $d(\gamma)$ et par suite $\delta(H) = H + \sigma(H) \subset H + \rho(\gamma)$; H et $\sigma(H)$ étant disjoints, comme le sont d'ailleurs H et $\rho(\gamma)$, on en conclut $\sigma(H) \subset \rho(\gamma)$. Donc $\phi[\sigma(H)] = 0$, d'après le second caractère $\phi[\rho(\gamma)] = 0$ des familles G régulières (18). Donc H est un noyau.

Tout point M de $\Delta(G)$ est un noyau.

Observons d'abord que, la mesure-ϕ étant additive et non négative, $\phi(M)$ existe et vaut zéro. En effet, quel que soit l'ensemble γ de G contenant M, $\bar{\phi}_e(M) \leq \phi(\gamma)$. Et, M étant dans $\Delta(G)$, la plus petite limite des nombres $\phi(\gamma)$ correspondants est zéro. Donc $\phi_e(M) \leq 0$, soit $\phi(M) = 0$.

La famille $G(M)$ est formée des γ de G contenant M. Pour chacun d'eux $M \subset \gamma \cdot \Delta$; $\delta(M) = M + \sigma(M)$, ensemble indéfiniment couvert par les γ de $G(M)$, est inclus dans $\delta(\gamma \cdot \Delta) = \gamma \cdot \Delta + \rho(\gamma)$. Donc

$$\phi[\delta(M)] = \phi[\sigma(M)] \leqq \phi(\gamma \cdot \Delta) + \phi[\rho(\gamma)] \leqq \phi(\gamma).$$

Cette relation étant vérifiée quel que soit γ contenant M, et M étant dans $\Delta(G)$, $\phi[\delta(M)] = 0$; M est un ensemble-noyau.

A deux ensembles-noyaux disjoints F et F' il ne correspond pas nécessairement un nombre positif α tel qu'aucun γ de G, de mesure $\phi(\gamma) < \alpha$, ne puisse être en même temps joint à F et à F'.

Les ensembles $F \cdot \sigma(F')$ et $F' \cdot \sigma(F)$ peuvent ne pas être vides tous les deux. En ce cas α n'existe certainement pas. Si ces deux ensembles sont vides, mais si $\sigma(F) \cdot \sigma(F')$ existe, il peut se faire qu'une suite d'ensembles γ, de mesure $\phi(\gamma)$ tendant vers 0, soient tous joints à F et à F' simultanément.

L'hypothèse que deux ensembles noyaux soient disjoints n'implique rien qui fasse songer à une sorte de distance mutuelle positive des deux ensembles.

32. *La somme d'un nombre fini d'ensembles-noyaux est un ensemble-noyau* (relativement à une famille régulière G).

Soient $F_1, \cdots, F_p$ les ensembles-noyaux donnés et $F = \Sigma F_i$; $G(F) = \Sigma G(F_i)$ (tout γ joint à F est joint à l'un des F_i, ce serait vrai même si les F_i n'étaient pas en nombre fini). $\delta(F) = \Sigma\delta(F_i)$; car un point M indéfiniment couvert par la réunion des $G(F_i)$ est indéfiniment couvert par l'une au moins de ces familles, parce qu'elles sont en nombre fini. Donc $\delta(F) = \Sigma F_i + \Sigma\sigma(F_i) = F + \sigma(F)$, donc $\sigma(F) \subset \Sigma\sigma(F_i)$ et $\phi[\sigma(F)] = 0$.

Si l'ensemble commun à une famille dénombrable d'ensembles-noyaux existe, c'est un ensemble-noyau (relativement à la famille G et à la métrique-ϕ).

Soit $H_1, H_2, \cdots, H_n, \cdots$ une suite d'ensembles-noyaux relativement à G et $H = \Pi H_n$ leur ensemble commun majeur. Les H_n sont mesurables-ϕ. La métrique étant borélienne, H est mesurable-ϕ, s'il n'est pas vide. Si $H_n = H + J_n$, H et J_n étant disjoints, l'ensemble commun aux J_n est vide, puisque H est l'ensemble majeur commun aux H_n. Formons $G(H)$. C'est l'ensemble des γ de G joints à H. Un tel γ est *a fortiori* joint à H_n, quel que soit n, puisque H_n contient H. Donc $G(H) \subset G(H_n)$ et $\delta(H) \subset \delta(H_n)$, soit:

$$H + \sigma(H) \subset H_n + \sigma(H_n) = H + J_n + \sigma(H_n).$$

H et $\sigma(H)$ sont disjoints, et il en est de même de H, J_n et $\sigma(H_n)$. Donc $\sigma(H) \subset J_n + \sigma(H_n)$.

$\sigma(H)$ est donc commun à tous les $J_n + \sigma(H_n)$. Un point M de $\sigma(H)$ ne peut pas être commun à tous les J_n, puisque ΠJ_n est vide. Donc M est commun à tous les $\sigma(H_n)$; $\sigma(H) \subset \sigma(H_n)$. Or, H étant un noyau, $\phi[\sigma(H_n)] = 0$. Donc $\phi[\sigma(H)] = 0$. H est relativement à G un ensemble-noyau.

En particulier si les H_n décroissent ($H_n \supset H_{n+1}$) et si $\phi(H_n)$ qui est non croissant a pour limite le nombre λ, ou bien $\lambda > 0$ et alors le noyau $H = \Pi H_n$ existe et a pour mesure λ; ou bien $\lambda = 0$ et $H = \Pi H_n$ peut ne pas exister. S'il existe, sa mesure-ϕ est 0.

33. La condition que la famille des ensembles-noyaux soit dénombrable est essentielle. Par exemple, dans l'espace linéaire U_1 de l'axe des x, et avec la métrique borélienne des longueurs, soit e_0 un ensemble situé sur l'intervalle $s_0(0, 1)$ et de longueur extérieure 1, de longueur intérieure 0; e_0 a la puissance du continu, et il en est de même de $\sigma \cdot e_0$, quel que soit le segment σ contenu dans s_0; cet ensemble $\sigma \cdot e_0$ a la longueur extérieure σ et la longueur intérieure 0.

Considérons une famille G d'ensembles γ, G se décomposant en familles $G_{k,h}$ ($k \geq 1$, $h = 1, \cdots, 2^k$) ainsi constituées: Tout ensemble γ de $G_{k,h}$ est le segment $\sigma_{k,h}$ $[(h-1)2^{-k} \leq x \leq h2^{-k})$ diminué d'un point de $e_0 \cdot \sigma_{k,h}$. Pour chaque système d'indices (k, h), les γ de $G_{k,h}$ sont en infinité ayant la puissance du continu. La famille $G = \sum_{k,h} G_{k,h}$ est composée d'ensembles γ ayant chacun une longueur déterminée. La famille G est régulière. Chacun des γ de G est un noyau.

$\Delta(G)$, ensemble indéfiniment couvert au sens des longueurs par les γ de G, est identique au segment $(0, 1)$. Si l'on considère la famille des γ de $G_{k,h}$, famille non dénombrable, composée d'ensembles-noyaux relativement à G, l'ensemble commun à tous les ensembles de cette famille est $\sigma_{k,h} - \sigma_{k,h} \cdot e_0$, qui est non mesurable, puisque $\sigma_{k,h} \cdot e_0$ est non mesurable. Cet ensemble commun n'étant pas mesurable, n'est pas un ensemble-noyau.

34. Relativement à la métrique-ϕ et à une famille régulière G d'ensembles γ, nous appelons *ensemble-enveloppe le complémentaire par rapport à* $\Delta = \Delta(G)$ *d'un ensemble-noyau.*

C'est là une première définition. En voici une seconde, équivalente à la première:

Relativement à la métrique-ϕ et à la famille régulière G d'ensembles γ, l'ensemble O mesurable-ϕ, inclus dans Δ, est un ensemble-enveloppe si la *famille $G'(O)$ des γ de G disjoints de $\Delta - O$ est telle que les points de O indéfiniment couverts par la famille $G - G'(O) = G(\Delta - O)$, complémentaire de $G'(O)$ par rapport à G, forment un ensemble $\tau(O)$ de mesure-ϕ nulle.*

Un ensemble O' satisfaisant à la première définition, donc complémentaire d'un ensemble noyau $F = \Delta - O'$ satisfait à la seconde.

En effet, la famille $G - G'(O') = G(\Delta - O')$ est $G(F)$. Posons $\tau(O') = \sigma(F)$. Les points de $O' = \Delta - F$ converts indéfiniment par $G(F)$ forment bien l'ensemble $\tau(O')$ et $\phi[\tau(O')] = \phi[\sigma(F)] = 0$.

Réciproquement, soit O vérifiant la seconde définition et $F' = \Delta - O$. Je dis que F' est noyau. En effet, $G(\Delta - O)$ est $G(F')$, et puisque par hypothèse l'ensemble $\tau(O)$ est celui des points $O = \Delta - F'$ couverts indéfiniment par $G(F')$, $\sigma(F')$ est identique à $\tau(O)$ et d'après l'hypothèse $\phi[\tau(O)] = 0$, on conclut $\phi[\sigma(F')] = 0$; F' est noyau.

Voici une remarque utilisée dans ma définition initiale des ensembles-enveloppes, donnée dans ma Note aux *Comptes Rendus* du 13 novembre 1950:

En vertu de notre seconde définition de l'ensemble-enveloppe O, si M est un point de $O - \tau(O)$, les γ de G contenant M et étrangers à $G'(O)$ ont leurs mesures $\phi(\gamma)$ bornées inférieurement par un nombre positif α, dépendant évidemment de M.

Car les γ de G étrangers à $G'(O)$ sont dans $G(\Delta - O)$; M étant dans $O - \tau(O)$ n'est pas indéfiniment couvert par la famille $G(\Delta - O)$, en sorte que les γ de $G(\Delta - O)$ contenant M vérifient bien la condition énoncée.

Un point de $\tau(O)$ est indéfiniment couvert par $G(\Delta - O)$. Il l'est ou ne l'est pas par $G'(O)$.

35. COROLLAIRES. 1°. *L'ensemble commun à un nombre fini d'ensembles enveloppes est un ensemble-enveloppe.*

2°. *La somme d'une infinité dénombrable d'ensembles-enveloppes est un ensemble-enveloppe.*

Si $O_1, O_2, \cdots, O_n, \cdots$ sont des enveloppes, les ensembles $F_n = \Delta - O_n$ sont des noyaux. $\Pi O_n = \Delta - \Sigma F_n$, $\Sigma O_n = \Delta - \Pi F_n$.

Si les O_n sont en nombre fini, ΣF_n est un ensemble-noyau; ΠO_n est une enveloppe. Que les O_n soient en nombre fini ou en infinité dénombrable, ΠF_n est un noyau, donc ΣO_n est une enveloppe.

L'exemple du n° 33 montre que, si les ensembles enveloppes étaient en infinité indénombrable, leur somme ne serait plus nécessairement une enveloppe.

3°. *Si F est un ensemble-noyau et O un ensemble enveloppe, $F - F \cdot O$ est noyau, $O - O \cdot F$ est enveloppe.*

Car $F - F \cdot O = F \cdot (\Delta - O)$ est l'intersection de deux noyaux; $O - O \cdot F = O \cdot (\Delta - F)$ est l'intersection de deux enveloppes.

4°. *Si F ensemble-noyau est inclus dans O, ensemble-enveloppe, l'ensemble Θ indéfiniment couvert par les γ de G joints à F et disjoints de $\Delta - O = f$ vérifie les inégalités (égalités non exclues)*:

$$F - F \cdot \tau(O) \subset \Theta \subset F + O \cdot \sigma(F).$$

La famille des γ joints à F [donc inclus dans $G(F)$] et disjoints de f [donc inclus dans $G'(O)$] est $G(F) \cdot G'(O) = g$.

Un point M de F étranger à $\tau(O)$ est dans $O - \tau(O)$; M est donc couvert indéfiniment par des ensembles γ de $G'(O)$. Ces γ contenant M sont dans $G(F)$. Ces γ sont dans g. Donc M est dans Θ et $\Theta \supset F - F \cdot \tau(O)$.

D'autre part, tout point de Θ doit être couvert indéfiniment par $G(F)$, donc il est dans $F + \sigma(F)$. Il n'est pas dans f, dont tous les γ de $G'(O)$, et *a fortiori* de g, sont disjoints; donc il est dans $O \cdot [F + \sigma(F)] = F + O \cdot \sigma(F)$. La seconde inclusion est établie.

$\Delta = \Delta(G)$ *est à la fois noyau et enveloppe par rapport à la famille régulière* G. Car $G(\Delta) = G'(\Delta) = G$; $\sigma(\Delta)$ et $\tau(\Delta)$ sont vides.

Dans le plan U_2, la fonction métrique $\phi(E)$ étant l'aire borélienne, et si la famille G est composée des carrés contenus dans un ensemble ouvert D, les noyaux F sont les ensembles fermés diminués d'un ensemble d'aire nulle (ou inexistant), les enveloppes sont les ensembles ouverts accrus d'un ensemble d'aire nulle.

36. *Nature antitopologique des espèces considérées.* Il est à remarquer qu'à l'inverse de l'ordre adopté en topologie commune, la notion d'ensemble-noyau, évoquant, mais très vaguement, celle d'ensemble fermé des espaces cartésiens, se définit directement, tandis que celle d'ensemble-enveloppe, qui rappellerait, de loin, celle d'ensemble ouvert, s'introduit par la voie de la complémentarité et se rattache à la première. En topologie générale, le caractère de l'ensemble fermé est intrinsèque. Il se conserve dans une extension de l'espace contenant l'ensemble, pourvu que la topologie reçoive l'extension correspondante. Au contraire, l'ensemble ouvert de l'espace initial cesse d'être ouvert dans l'extension. Quel que soit l'espace variable contenant un ensemble fermé immuable, le complémentaire de celui-ci par rapport à l'espace variable est toujours un ensemble ouvert, mais dont la nature ne cesse pas de changer avec l'espace. Au contraire l'ensemble fermé garde constamment son caractère.

Une topologie qui donnerait la première place aux caractères intrinsèques des ensembles ne se fonderait pas sur les ensembles ouverts.

37. L'ensemble $\Delta = \Delta(G)$ est défini comme étant l'ensemble des points de l'espace U indéfiniment couverts par les ensembles γ de la famille G incluse dans U. Mais un point M de Δ n'est caractérisé que par la famille $G(M)$ des γ de G contenant ce point M, $G(M)$ couvrant indéfiniment M. Aussi cette famille $G(M)$ caractérise-t-elle, solidairement avec M, la totalité des autres points de $\Delta[G(M)] = \delta(M) = M + \sigma(M)$, à savoir la totalité des points N de $\sigma(M)$ qui, distincts de M, sont indéfiniment couverts, avec M, par $G(M)$.

Ecrivons $N \sim M$, cela signifiant que N est dans $\sigma(M)$. Mais, puisque parmi les γ contenant M, il en existe une infinité contenant N, donc appartenant à $G(N)$, et dont la mesure $\phi(\gamma)$ tend vers zéro, il en résulte que M est dans $\sigma(N)$. Donc la relation $N \sim M$ entraîne $M \sim N$. Cette liaison, exprimée par le signe $\sim$, est réflexive, selon le langage usuel d'aujourd'hui. Nous disons que M et N sont *indiscernables* dans Δ et relativement à la famille G.

Parmi les points de $\sigma(M)$ distincts de N, il peut s'en trouver un, soit P, tel que les γ de $G(M)$ contenant P ne couvrent pas indéfiniment N, tandis qu'ils couvrent indéfiniment P, puisque P est par hypothèse dans $\sigma(M)$. En ce cas, il pourra se faire que $G(P)$ ne couvre pas indéfiniment N, ni corrélativement que $G(N)$ ne couvre pas indéfiniment P, $G(N)$ et $G(P)$ couvrant indéfiniment l'un et l'autre le point M. Alors les deux couples de relations réversibles $(N \sim M, M \sim N)$ et $(P \sim M, M \sim P)$ n'entraînent pas $(P \sim N, N \sim P)$. Tout cela est évidemment très éloigné des conditions d'une topologie.

Nous dirons que M et N sont *totalement indiscernables par* G si, ces deux points étant dans $\Delta = \Delta(G)$, les deux ensembles $\delta(M)$ et $\delta(N)$ sont identiques et si de plus les γ de $G(M)$ ne contenant pas N et les γ de $G(N)$ ne contenant pas M ont leurs mesures $\phi(\gamma)$ bornées inférieurement par un nombre positif: $\delta(M) = \delta(N) = \Delta[G(M) \cdot G(N)]$.

38. Relativement à la métrique-ϕ dans l'espace U et à la famille régulière G d'ensembles γ, appelons *clos métrique* d'un ensemble H contenu dans $\Delta = \Delta(G)$ l'ensemble $\delta(H) = \Delta[G(H)]$ indéfiniment couvert par les γ de G joints à H. L'ensemble $\delta(H)$ qui contient H, joue par rapport à H un role qui fait songer à celui de l'ensemble de fermeture $\bar{E}$ d'un ensemble E

en topologie. Mais nous trouvons une opposition capitale entre les deux notions dans le fait que $\bar{\bar{E}} = \bar{E}$ (un ensemble de fermeture est identique à son propre ensemble de fermeture), tandis qu'il n'y a aucune raison logique pour que $\delta[\delta(H)]$, le clos métrique de $\delta(H)$, l'ensemble indéfiniment couvert par les γ de G joints au clos métrique $\delta(H)$ de H, soit identique à $\delta(H)$.

Notons que, d'après $\delta(H) = H + \sigma(H)$,

$$\delta[\delta(H)] = \delta(H) + \delta[\sigma(H)] = \delta(H) + \sigma(H) + \sigma[\sigma(H)] = \delta(H) + \sigma[\sigma(H)].$$

Le clos métrique $\delta(F)$ d'un noyau F a même mesure que celui-ci. Mais on ne voit pas que ce clos métrique soit nécessairement un noyau. De ce que $\phi[\sigma(F)] = 0$, il ne paraît y avoir aucune nécessité pour que la mesure-ϕ de $\sigma[\sigma(F)]$ soit elle aussi nulle. S'il n'en est pas ainsi, $\delta(F)$ n'est pas un noyau, bien que F le soit.

39. A la lumière de l'étude précédente, revenons aux propositions des n$^{\text{os}}$ 22 et 28.

Soit F un ensemble noyau de la famille régulière G formée d'ensembles γ. La totalité des γ de G joints à F constitue la famille $G(F)$ couvrant indéfiniment $\delta(F) = F + \sigma(F)$, F et $\sigma(F)$ étant disjoints avec $\phi[\sigma(F)] = 0$.

Soit M un point de F. Les épaisseurs diverses (supérieure, inférieure, éventuellement exacte) de F au point M (22) relatives à la famille G sont les limites du rapport $\phi(F \cdot \gamma)/\phi(\gamma)$, l'ensemble γ de G contenant M et $\phi(\gamma)$ tendant vers 0. Mais, puisque G ne renferme aucun ensemble γ étranger à $G(M)$ et contenant M, il y a identité entre les épaisseurs de F en son point M relativement à G et relativement à $G(F)$. Le théorème 22 sur l'épaisseur de l'ensemble $\Delta = \Delta(G)$ relativement à G est vrai de l'épaisseur de l'ensemble $\delta(F)$ relativement à $G(F)$, puisque $\delta(F) = \Delta[G(F)]$. En conséquence, l'épaisseur de $\delta(F)$ relativement à $G(F)$ est 1 sur une plénitude-ϕ de $\delta(F)$. Soit H cette plénitude de $\delta(F)$; H est inclus dans $\delta(F)$ et $\phi(H) = \phi[\delta(F)] = \phi(F)$, puisque $\phi[\sigma(F)] = 0$, F étant noyau.

D'après $H \subset F + \sigma(F)$, $H = H \cdot F + H \cdot \sigma(F)$, et, avec $\phi[H \cdot \sigma] = 0$, $\phi(H \cdot F) = \phi(H) = \phi(F)$; $H' = H \cdot F$ est une plénitude-ϕ de F. En tout point M de H', l'épaisseur de $\delta(F) = F + \sigma(F)$ relative à $G(F)$ est 1.

D'autre part, pour tout ensemble γ, $\gamma \cdot \delta(F) = \gamma \cdot F + \gamma \cdot \sigma(F)$. Donc $\gamma \cdot \delta(F)$ a même mesure que $\gamma \cdot F$; donc l'épaisseur de $\delta(F)$ et celles de F relativement à $G(F)$ sont égales en tout point. En conséquence, l'épaisseur de F par rapport à $G(F)$ est 1 en tout point de H'. Mais, comme nous l'avons vu, l'épaisseur de F par rapport à $G(F)$ ou par rapport à G est la même en tout point de F. Donc, *relativement à la famille G, l'épaisseur de* *l'ensemble-noyau F est* 1 *sur une plénitude-ϕ de F. Je dis que, sur une plénitude de l'ensemble-enveloppe $O = \Delta - F$, l'épaisseur de F est* 0.

En effet, $\tau(O)$ étant $\sigma(F)$, soit M un point de $O - \tau(O) = \Delta - \delta(F)$; $O - \tau(O)$ a même mesure que O; M étranger à $\delta(F)$ n'est pas couvert indéfiniment par $G(F)$. Les γ de G contenant M et joints à F ont (33) leur mesure $\phi(\gamma)$ bornée inférieurement par un nombre positif α, dépendant de M et de F). Si γ contient M et si $\phi(\gamma) < \alpha$, $\phi(F \cdot \gamma) = 0$, puisque $F \cdot \gamma$ est vide. Donc $\lim \phi(\gamma \cdot F)/\phi(\gamma) = 0$ quand γ contenant M a sa mesure $\phi(\gamma)$ tendant vers 0. L'épaisseur de F en tout point M de $O - \tau(O)$ est nulle. Or d'après $\phi[\tau(O)] = \phi[\sigma(F)] = 0$, $O - \tau(O)$ est une plénitude-ϕ de $O = \Delta - F$.

40. Au n° 26 nous avons considéré un ensemble H tel que, avec les notations de la presente étude complémentaire, l'ensemble $\Delta[G(H)] = \delta(H) = H + \sigma(H)$ ait même mesure que H. Un tel ensemble est donc, dans le cas de toute famille régulière G, un ensemble-noyau. Sans autre hypothèse sur G que sa régularité, en-dehors des points isolés de Δ, des ensembles $\gamma \cdot \Delta$, et des sommes finies de ces variétés, nous ne sommes assurés de l'existence que d'un seul ensemble-noyau relativement à G, à savoir $\Delta = \Delta(G)$ lui-même.

Nous avons montré, et il n'y a rien à retrancher à la démonstration, que, si la fonction d'ensemble $\psi(E)$ est définie, bornée, et métriquement continue dans le champ $\Delta(G)$ (25), les hypothèses indépendantes $\Lambda_i(\psi, M, G) < A$, $\Lambda_s(\psi, M, G) > B$ vérifiées en tout point M d'un ensemble noyau H (A et B ne dépendent pas de M), entraînent respectivement $\psi(H) < A\phi(H)$, $\psi(H) > B\phi(H)$.

$\Lambda_i(\psi, M, G)$, $\Lambda_s(\psi, M, G)$ sont, au point M, les nombres dérivés inférieur et supérieur, relativement à la famille G, de la fonction d'ensemble $\psi(E)$ (23). On verrait comme plus haut qu'en un point M de H, les nombres dérivés par rapport à G et par rapport à $G(H)$ sont les mêmes. Notons que, d'après $\phi[\sigma(H)] = 0$ et la fonction d'ensemble $\psi(E)$ étant métriquement continue, quel que soit γ dans G: $\psi[\gamma \cdot \delta(H)] = \psi(\gamma \cdot H)$, (comme dans le cas particulier où la fonction ψ est la mesure ϕ).

41. Enfin, venons -en à la définition rectifiée des familles G d'ensembles γ de l'espace U, parfaitement régulières relativement à la métrique-ϕ adoptée dans U.

La famille G sera dite *parfaitement régulière,* si, ϵ étant un nombre positif indépendant, tout ensemble H mesurable-ϕ inclus dans Δ contient un ensemble-noyau F et est inclus dans un ensemble-enveloppe O, de façon que $\phi(H) - \phi(F)$ et $\phi(O) - \phi(H)$ soient inférieurs à ϵ.

Tout ensemble-noyau F étant mesurable-ϕ, F est contenu dans un ensemble-enveloppe O vérifiant $\phi(O) < \phi(F) + \epsilon$. Et tout ensemble enveloppe O' de mesure-ϕ supérieure à ϵ contient un ensemble-noyau F' avec $\phi(F') > \phi(O') - \epsilon$.

Ces propriétés ne sont nullement certaines avec des familles G simplement, mais non parfaitement, régulières.

Quand la famille G est parfaitement régulière, il est possible, relativement à tout ensemble H mesurable-ϕ, inclus dans Δ, de trouver une famille $g(H)$ d'ensembles γ de G, telle que l'ensemble $\omega(H) = \Delta[g(H)]$, formé des points indéfiniment couverts par les γ de $g(H)$ ne diffère de H que par un ensemble de mesure-ϕ nulle.

La suite numérique ϵ_n tendant vers 0, soient F_n noyau, O_n enveloppe, deux ensembles vérifiant $F_n \subset H \subset O_n$, $\phi(O_n) - \phi(F_n) < \epsilon_n$. D'après les théorèmes du n° 32 et 35, on peut supposer F_n croissant (non décroissant) et O_n décroissant (non croissant). Posons $f_n = \Delta - O_n$; f_n est noyau et croît avec n. Soit g_n la famille $G(F_n) \cdot G'(O_n)$ et Θ_n l'ensemble $\Delta(g_n)$. Nous avons vu (35) que

$$F_n - F_n \cdot \tau(O_n) \subset \Theta_n \subset F_n + O_n \cdot \sigma(F_n).$$

Nous posons $g(H) = \Sigma g_n$.

Désignons par ρ_p la famille $\sum\limits_{n>p} g_n$; $g(H) = g_1 + g_2 + \cdots + g_p + \rho_p$.

Pour $n > p$, $G'(O_n)$ disjoint de f_n est *a fortiori* disjoint de $f_p \subset f_n$. *A fortiori* g_n est-il disjoint de f_p, donc inclus dans $G'(O_p)$. En conséquence, $\Delta(\rho_p) \subset O_p$.

Soit $\omega_n = \Theta_1 + \Theta_2 + \cdots + \Theta_n$.

$\omega(H) = \omega_p + \Delta(\rho_p)$. Donc, d'après $F_1 + F_2 + \cdots + F_p = F_p$ et $F_p + O_p = O_p$,

$$\omega(H) \supset F_p - \sum_{n \leqq p} F_n \cdot \tau(O_n) \quad \text{et} \quad \omega(H) \subset \sum_{n \leqq p} O_n \cdot \sigma(F_n) + O_p.$$

Quand p croît, F_p tend en croissant vers $\sum\limits_{n \geqq 1} F_n = H - \lambda$, avec $\phi(\lambda) = 0$; O_p tend en décroissant vers $\prod\limits_{n \geqq 1} O_p = H + \mu$, avec $\phi(\mu) = 0$.

Donc

$$H - \lambda - \sum_{n \geqq 1} F_n \cdot \tau(O_n) \subset \omega(H) \subset H + \mu + \sum_{n \geqq 1} O_n \cdot \sigma(F_n).$$

Les $\phi(\tau)$ et $\phi(\sigma)$ étant nuls. $H - H \cdot \omega(H)$ et $\omega(H) - H \cdot \omega(H)$ ont l'un et l'autre la mesure-ϕ nulle, $\omega(H)$ étant $\Delta[g(H)]$.

42. THÉORÈME. *Relativement à une famille G parfaitement régulière, l'épaisseur de tout ensemble H mesurable-ϕ inclus dans $\Delta(G)$ est 1 sur une plénitude-ϕ de H et 0 sur une plénitude-ϕ de $\Delta - H$.*

Car, quel que soit l'ensemble F inclus dans H, l'épaisseur de F en un quelconque de ses points est au plus égale à l'épaisseur de H au même point. Si F est un ensemble-noyau, son épaisseur relativement à G est 1 sur une plénitude-ϕ de F. Or $\phi(F)$ peut être supposé plus grand que $\phi(H) - \epsilon$. Donc, l'ensemble $J(H)$ des points de H où l'épaisseur $\eta(H, M)$ est 1, a une mesure supérieure à $\phi(H) - \epsilon$ quel que soit $\epsilon > 0$. Donc, $J(H)$ a pour mesure $\phi(H)$. C'est une plénitude-ϕ de H.

$\Delta - H$ étant mesurable-ϕ a l'épaisseur 1 sur une plénitude-ϕ de lui même. Donc, sur une plénitude de $\Delta - H$, H a l'épaisseur 0. Car évidemment $\eta_i(H) + \eta_s(\Delta - H) = 1$, $\eta_s(H) + \eta_i(\Delta - H) = 1$, en tout point M de Δ.

43. THÉORÈME. *Relativement à la famille G parfaitement régulière selon une métrique-ϕ, toute fonction d'ensemble $\psi(E)$ définie, bornée, métriquement continue dans le champ $\Delta = \Delta(G)$ a une dérivée finie sur une plénitude-ϕ de Δ.*

Sans quoi, il existerait (26) deux nombres A, B avec $A < B$ et un ensemble H de mesure-ϕ positive tel qu'en tout point M de H, $\Lambda_i(\psi, M, G) < A$, $\Lambda_s(\psi, M, G) > B$. Si H n'est pas un noyau relativement à G, et si G est simplement, mais non parfaitement régulière, nous ne pouvons rien tirer de ces conclusions. Car $\phi[\sigma(H)] > 0$. Par contre, G étant parfaitement régulière et, d'après $\phi(H) > 0$, H contient un ensemble-noyau F de mesure-ϕ positive. Or, les deux hypothèses précédentes, vérifiées l'une et l'autre sur le noyau F sont incompatibles, puisqu'elles entraîneraient $\psi(F) < A\phi(F)$ et $\psi(F) > B\phi(F)$.

OBSERVATION. *Si, la famille G étant parfaitement régulière, en tout point M d'un ensemble H mesurable-ϕ inclus dans $\Delta = \Delta(G)$ la fonction d'ensemble $\psi(E)$, définie, bornée et métriquement continue dans le champ Δ vérifie la condition $\Lambda_i(\psi, M, G) < A$* [ou la condition $\Lambda_s(\psi, M, G) > B$], *$A$ (ou B) étant un nombre indépendant de M variant sur H, on en conclut $\psi(H) < A\phi(H)$* [ou $\psi(H) > B\phi(H)$].

Traitons l'hypothèse $\Lambda_i < A$ sur H.

H étant mesurable-ϕ et inclus dans Δ contient un ensemble-noyau F et si $J = H - F$, on peut supposer $\phi(J) < \epsilon$. L'inégalité $\Lambda_i < A$ est vérifiée sur F; F étant noyau, on en conclut (40) $\psi(F) < A\phi(F)$. Donc

$$\psi(H) = \psi(F) + \psi(J) < A[\phi(H) - \phi(J)] + \psi(J).$$

On peut faire tendre $\phi(J)$ vers 0, par une succession de choix de F; $\psi(E)$ étant métriquement continue, $\psi(J)$ tend vers 0 avec $\phi(J)$. Donc $\psi(H) \leqq A\phi(H)$. D'ailleurs le signe d'égalité peut être supprimé. Car si, en tout point de H, $\Lambda_i < A$, il existe un α positif tel que l'ensemble des points de H où $\Lambda_i < A - \alpha$ a pour mesure-ϕ un nombre positif λ, sans quoi $\Lambda_i \geqq A$ sur une plénitude-ϕ de H. Donc

$$\psi(H) \leqq A[\phi(H) - \lambda] + (A - \alpha)\lambda < A\phi(H).$$

Pareillement l'hypothèse $\Lambda_s > B$ en tout point de H entraîne

$$\psi(H) > B\phi(H).$$

44. Théorème. *Si, relativement à une métrique-ϕ et à une famille G d'ensembles γ parfaitement régulière, la fonction d'ensemble $\psi(E)$ est définie, bornée et métriquement continue dans le champ $\Delta = \Delta(G)$, la fonction $\Lambda(M)$ égale à la dérivée $\Lambda(\psi, M, G)$ aux points de Δ où cette dérivée existe et est finie, et à zéro sur l'ensemble complémentaire R, $\Lambda(M)$ est sommable-ϕ sur Λ et, quel que soit E mesurable-ϕ inclus dans Δ, $\psi(E)$ est égal à $\mathfrak{J}_L[\Lambda(M)d\phi, E]$, intégrale lebesguienne de $\Lambda(M)d\phi$ sur E.*

La démonstration est du type le plus classique. Puisque $\psi(E)$ a relativement à G une dérivée finie sur une plénitude-ϕ de Δ, $\phi(R) = 0$ et en conséquence $\psi(R) = 0$.

Soient E^+ et E^- les ensembles des points de E où respectivement $\Lambda(M) > 0$ et $\Lambda(M) < 0$; $\psi(E) = \psi(E^+) + \psi(E^-)$. Car, sur $E' = E - E^+ - E^-$ (E' contient $E \cdot R$), $\Lambda(M) = 0$ et par suite $\psi(E') = 0$, indépendamment de la mesure-ϕ de E' [Sur E', $-\epsilon < \Lambda(M) < \epsilon$ quel que soit $\epsilon > 0$ indépendant, donc $-\epsilon\phi(E') < \psi(E') < \epsilon\phi(E')$].

Divisons l'intervalle $(0, +\infty)$ par une suite de nombres croissant indéfiniment l_i $(i = 0, 1, 2, \cdots)$ avec $l_0 = 0$, $l_i - l_{i-1} < \omega$, ω étant indépendant de i $(\geqq 1)$. Soit E_i l'ensemble des points de E (de E^+) où $l_{i-1} < \Lambda(M) \leqq l_i$; $E^+ = \Sigma E_i$ $(i \geqq 1)$ et $l_{i-1}\phi(E_i) < \psi(E_i) \leqq l_i\phi(E_i)$ (43), donc $\psi(E_i) = \lambda_i\phi(E_i)$, λ_i étant un certain nombre vérifiant $l_{i-1} < \lambda_i \leqq l_i$. Les $\psi(E_i)$ étant positifs, et la fonction $\psi(E)$ étant bornée dans le champ Δ, la série $\Sigma\psi(E_i)$ converge et $\psi(E^+) = \Sigma\psi(E_i) = \Sigma\lambda_i\phi(E_i)$. Donc $\Lambda(M)$ est sommable-ϕ sur E^+ et $\psi(E^+) = \mathfrak{J}_L[\Lambda(M)d\phi, E^+]$.

De même $\Lambda(M)$ est sommable-ϕ sur E^- et $\psi(E^-) = \Im_L[\Lambda(M)\,d\phi, E^+]$. $\Lambda(M)$ étant sommable-ϕ sur E^+, sur E^- et aussi sur E' où $\Lambda(M) = 0$, $\Lambda(M)$ est sommable-ϕ sur E et $\psi(E) = \Im_L[\Lambda(M)\,d\phi, E]$.

E peut être identique à Δ. Donc $\Lambda(M)$ est sommable-ϕ sur Δ.

45. Théorème (réciproque ou précédent). *Si, la famille G étant parfaitement régulière relativement à la métrique-ϕ, la fonction de point $f(M)$ est définie et sommable-ϕ sur l'ensemble $\Delta = \Delta(G)$, l'intégrale lebesguienne $\Im_L[f(M)\,d\phi, E] = \psi(E)$, où E est un ensemble quelconque mesurable-ϕ, inclus dans Δ, est une fonction d'ensemble définie, bornée et métriquement continue dans le champ Δ, et la dérivée de $\psi(E)$ relativement à G est $f(M)$ sur une plénitude-ϕ de Δ.*

On suppose la fonction $f(M)$ finie en tout point de Δ.

La fonction d'ensemble $\psi(E)$ est définie dans le champ Δ, au sens donné à cette expression (23); $\psi(E)$ est bornée. Car, Δ^+ étant l'ensemble $f > 0$ et Δ^- l'ensemble $f < 0$,

$$|\psi(E)| < \Im_L[|f(M)|\,d\phi, \Delta] = \Im_L[f(M)d\phi, \Delta^+] - \Im_L[f(M)d\phi, \Delta^-] < \infty.$$

Enfin $\psi(E)$ est métriquement continue dans Δ. Car, si $\phi(E)$ tend vers 0, $\psi(E) = \Im_L[f(M)\,d\phi, E]$ tend vers 0. Car f étant sommable-ϕ, cette dernière intégrale tend vers 0.

En conséquence, relativement à G et sur un ensemble $\Delta - R$ qui est une plénitude-ϕ de Δ, $\psi(E)$ a une dérivée finie. $\Lambda(M)$ étant cette dérivée sur $\Delta - R$ et 0 sur R, nous venons de montrer que $\Lambda(M)$ est sommable-ϕ sur Δ et que $\psi(E) = \Im_L[\Lambda(M)\,d\phi, E]$ quel que soit E mesurable-ϕ dans Δ. Donce si $g(M) = \Lambda(M) - f(M)$, $\Im_L[g(M)\,d\phi, E] = 0$ quel que soit E mesurable-ϕ avec $E \subset \Delta$. En conséquence, si petit que soit $\epsilon > 0$, les ensembles $g(M) > \epsilon$ et $g(M) < -\epsilon$ ont la mesure-ϕ nulle. Donc $g(M) = 0$ et $f(M) = \Lambda(\psi, M, G)$ sur une plénitude-ϕ de Δ.

Il ressort de tout ce qui précède que les propriétés des fonctions absolument continues d'une variable réelle s'étendent aux fonctions d'ensembles métriquement continues relatives aux familles parfaitement régulières, dans un espace et avec une métrique borélienne entièrement quelconques.

Paris, France.

ON NORMAL OPERATORS IN HILBERT SPACE.*

By C. R. PUTNAM.

---

1. For definitions and terminology used in this paper, see [3], especially pp. 27 ff.

The following Lemma is a slight generalization of a result obtained recently by Fuglede [1] (cf. also [2]) concerning normal operators:

LEMMA. *Let N_1 and N_2 be two normal operators in a Hilbert space with spectral resolutions*

$$N_j = \int z\, dK_z^j \qquad (z = x + iy;\ j = 1, 2). \tag{1}$$

If A is any bounded operator satisfying

$$AN_1 \subseteqq N_2 A, \tag{2}$$

then

$$AK_z^1 = K_z^2 A, \text{ for all } z, \tag{3}$$

and hence,

$$AN_1^* \subseteqq N_2^* A. \tag{4}$$

If $N_1 = N_2 = N$, the Lemma becomes exactly the Theorem of Fuglede.

The Lemma will be instrumental in proving a theorem concerning the equivalence of bounded normal operators in Hilbert space. The theorem in question is the following:

THEOREM I. *Let N_1 and N_2 denote two bounded normal operators for which there exists a non-singular bounded operator T such that*

$$N_2 = TN_1T^{-1}. \tag{5}$$

Then N_1 and N_2 are unitarily equivalent, that is, there exists a unitary operator U such that

$$N_2 = UN_1U^*. \tag{6}$$

In case both operators N_1 and N_2 are Hermitian or both are unitary, Theorem I is known. For the Hermitian case, the result is apparently due to Toeplitz, while the Theorem, in case both N_1 and N_2 are unitary, is due

---

* Received September 23, 1950.

to Wintner [7], p. 149. He has pointed out that a difficulty arises if one attempts to transcribe the proofs for these known cases to the present case, where N_1 and N_2 are supposed only normal ([8], pp. 139-140), and that the Theorem is surely true if N_1 and N_2 are finite normal matrices ([7], p. 149).

2. *Proof of the Lemma.* The proof of the Lemma will be given by indicating the appropriate modifications of the argument used by Fuglede [1]. Let α denote an arbitrary Borel set in the complex plane and define the operators $K^1(\alpha)$ and $K^2(\alpha)$ by

$$K^j(\alpha) = \int_\alpha dK_z^j, \qquad j = 1, 2.$$

Suppose now that α and β are disjoint squares of the lattice considered in [1] and let f denote an arbitrary element of $\mathfrak{D}_{N_1}$. Since A is bounded, the set $\mathfrak{D}_{N_1}$ is contained in the set $\mathfrak{D}_{AN_1}$, and, since $K^1(\alpha)f$ belongs to $\mathfrak{D}_{N_1}$, relation (2) implies

$$(7) \qquad AN_1K^1(\alpha)f = N_2AK^1(\alpha)f.$$

An application of the bounded operator $K^2(\beta)$ to each side of (7) and a utilization of the spectral resolutions (1) show that

$$(8) \qquad K^2(\beta)A\int_\alpha z\, dK_z^1 f = \int_\beta z\, dK_z^2\, AK^1(\alpha)f,$$

for all f in $\mathfrak{D}_{N_1}$. An argument parallel to that used in [1] shows that (8) implies

$$(9) \qquad K^2(\beta)AK^1(\alpha) = 0;$$

the Lemma follows easily from (9), cf. [1], and the proof is complete.

3. *Proof of Theorem* I. If λ denotes an arbitrary complex number and if I is the identity operator, then (5) holds if and only if $N_2 + \lambda I = T(N_1 + \lambda I)T^{-1}$. Since $N_j + \lambda I$ ($j = 1$ and 2) are normal, it is clear that it may be assumed that N_1 and N_2 are non-singular. This assumption will be convenient later in the proof.

For $j = 1, 2$, let $N_j = P_jU_j$ denote the polar decomposition of N_j, so that P_j is Hermitian positive-definite, U_j is unitary, and $P_jU_j = U_jP_j$. Equation (5) can be expressed as

$$(10) \qquad TN_1 = N_2T \qquad \text{or} \qquad TP_1U_1 = P_2U_2T,$$

and consequently, by the Lemma,

$$(11) \qquad TN_1^* = N_2^*T \qquad \text{or} \qquad TU_1^*P_1 = U_2^*P_2T.$$

The following set of equations is an obvious consequence of (10) and (11):

$$(12)\qquad TP_1{}^2 = TP_1U_1U_1{}^*P_1 = P_2U_2TU_1{}^*P_1 = P_2U_2U_2{}^*P_2T = P_2{}^2T.$$

Hence,

$$(13)\qquad TP_1 = P_2T;$$

this equation follows from (12) and from the fact that P_j, as the unique positive definite square root of $P_j{}^2$, can be expressed as a power series in $P_j{}^2$. (Relation (13) can also be obtained as a consequence of the Lemma.) Hence, by (10) and (13), $P_2U_2T = TP_1U_1 = P_2TU_1$, and therefore, since P_2 is non-singular,

$$(14)\qquad TU_1 = U_2T$$

If $T = PU$ denotes the polar decomposition of T, it follows from (13) and (14) that $UP_1 = P_2U$ and $UU_1 = U_2U$. In the latter (unitary) case, this result is given in [7], p. 149; the proof in the Hermitian case is similar. In fact, (13) implies

$$(15)\qquad PUP_1 = P_2PU,$$

and consequently, on taking the transposed conjugate of each side of (15), $P_1U^*P = U^*PP_2$. Multiplication on the left of each side of the last equation by PU and the use of (15) yield $P_2P^2 = P^2P_2$ and hence, as above, $P_2P = PP_2$. It follows from (15), that $PUP_1 = PP_2U$ and therefore $UP_1 = P_2U$, which was to be shown. Hence,

$$N_2 = P_2U_2 = UP_1U^*UU_1U^* = UP_1U_1U^* = UN_1U^*$$

that is, (6). This completes the proof of Theorem I.

4. The following theorem will be derived as a consequence of Theorem III (see below):

THEOREM II. *Let A denote an arbitrary bounded matrix and let B denote a bounded normal matrix. If* (i) $AB - BA = C$ *and if* (ii) $CB = BC$, *then* $C = 0$.

Even without the assumption (ii), the relation g. l. b. $| x^*Cx | = 0$, where x denotes a vector on the Hilbert sphere $\| x \| = 1$, is valid; cf. [4]. If, in addition, (ii) is assumed, it is clear that Theorem II greatly refines the assertion of the preceding sentence.

That the conclusion of Theorem II is false if B is not required to be normal is easily seen by verifying that the matrices

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

satisfy (i) and (ii), although $C \neq 0$.

The following theorem will be proved in Section **5**, by using Fuglede's theorem together with an adaptation of his methods:

THEOREM III. *In a Hilbert space, let B be a normal operator, with a spectral resolution*

$$B = \int z \, dK_z, \tag{16}$$

and let A, C be arbitrary operators satisfying

$$AB \subseteqq BA + C, \tag{17}$$

(i) *If, in addiaion to* (17), *C satisfies*

$$C \text{ is bounded and } \neq 0 \tag{18}$$

and

$$CB \subseteqq BC, \tag{19}$$

then A is unbounded. Furthermore, B cannot have an eigenfunction f belonging to $\mathfrak{D}_A$ and satisfying $Cf \neq 0$.

(ii) *If, in addition to* (17), *C satisfies $C = \lambda I + \mu B \neq 0$, where I is the identity operator and λ, μ denote complex constants, then A is unbounded. Furthermore, B cannot have an eigenfunction which belongs to $\mathfrak{D}_A$ and is associated with an eigenvalue z_0 satisfying $\lambda + \mu z_0 \neq 0$.*

Remark. If f is an eigenfunction of B belonging to $\mathfrak{D}_A$, it is clear that $Bf \subset \mathfrak{D}_A$ and hence, by (17), $f \subset \mathfrak{D}_C$.

Theorem II is an obvious consequence of part (i) of Theorem III.

It is known that if

$$AB = BA + \lambda I, \qquad \lambda \neq 0, \tag{20}$$

then A and B cannot both be bounded Hermitian operators; [9] or [5]. It follows from (ii) of Theorem III above that this result can be strengthened to: if A and B are self-adjoint ([3], p. 34) and satisfy (20) in the operator sense, then neither A nor B can be bounded. This is clear from the fact that the validity of (20) in the operator sense implies $AB \subseteqq BA + \lambda I$ and $BA \subseteqq AB - \lambda I$. In addition, it follows from (ii) above, since the number μ occurring there is zero, that neither A nor B can possess an eigenfunction belonging to the domain of the other. Rellich [5] has obtained results, which are rather explicit as far as the nature of the operators A and B is concerned, under different assumptions; namely, where (20) holds on a certain dense set and where A and B are merely symmetric (as distinguished from the more restrictive requirement of being self-adjoint) but are subject to certain other conditions.

Remark. It may be mentioned that, whether or not A or B is Hermitian, still (20) implies that A and B cannot *both* be bounded; cf. [6] and also the first paragraph of [4], wherein is cited a reference to [9].

5. *Proof of* (i) *of Theorem* III. Suppose, if possible, that A is bounded. For any Borel set α of the complex plane, define $K(\alpha) = \int_\alpha dK_z$. If f is in the domain of B (hence of AB) and if α and β have the same significance as in Section **2**, it follows that

$$(21) \qquad K(\beta)A\int_\alpha z dK_z f = \int_\beta z dK_z\, AK(\alpha)f + K(\beta)CK(\alpha)f.$$

Assumption (19) of Theorem III and an application of Fuglede's theorem (specifically the Lemma of Section **1** in which $N_1 = N_2$ and A become identified with the present B and C respectively), imply that $CK(\alpha) = K(\alpha)C$. Hence the second term on the right of (21) becomes $K(\beta)K(\alpha)Cf$. Since α and β are disjoint, it follows from the properties of the projection operators that $K(\beta)K(\alpha) = 0$; so that (8), in which the superscripts are omitted, is valid for all f in the domain of B. As in Section **2**, this leads to (9) and hence to $AK_z = K_zA$ for all z; cf. (3). The last equality implies however that $AB \subseteqq BA$ (cf. [1], p. 38, end of section 2). Since A is bounded, $\mathfrak{D}_{AB} = \mathfrak{D}_B$. Thus, the relations $AB \subseteqq BA$ and (17) imply $Cf = 0$ for all f in $\mathfrak{D}_B$, which is dense in the containing Hilbert space, $\mathfrak{H}$. Thus, since C is bounded, $Cf = 0$ for all $f \subset \mathfrak{H}$ $(= \mathfrak{D}_C)$. That is, $C = 0$, in contradiction with assumption (18) of Theorem III. Hence, A is unbounded and the proof of the first statement of Theorem III is complete.

Suppose, if possible, that z_0 is an eigenvalue of B with an eigenfunction f belonging to $\mathfrak{D}_A$ such that $Cf \neq 0$. The symbol $K(z_0)$, where z_0 is a Borel set consisting of one point, has been defined above (cf. also [3], p. 53). An application of the operator equation (17) to f (cf. the Remark following the statement of Theorem III) and a subsequent application of the bounded operator $K(z_0)$ to the resulting functional equation yield

$$(22) \qquad K(z_0)ABf = K(z_0)BAf + K(z_0)Cf.$$

In virtue of $K(z_0)C = CK(z_0)$, which follows from the theorem of Fuglede, and of the equations

$$Bf = z_0f, \qquad K(z_0)B = z_0K(z_0), \qquad K(z_0)f = f,$$

(cf. [3], p. 53), relations (22) implies that

$$(23) \qquad z_0K(z_0)Af = z_0K(z_0)Af + Cf,$$

in contradiction to the assumption $Cf \neq 0$. Thus, the supposition at the beginning of this paragraph is untenable and the proof of part (i) of Theorem III is complete.

Proof of (ii) *in Theorem* III. It is clear from the assumption of (ii) that C is normal and hence $K(\alpha)C \subseteqq CK(\alpha)$ for an arbitrary Borel set α. By a proof essentially identical with that given in the first paragraph of this section, it follows that A is unbounded. (In the present case, it should be noted that the assumption that A is bounded again leads to the conclusion that $Cf = 0$ for all $f \subset \mathfrak{D}_B$ $(= \mathfrak{D}_C)$. Since C is normal, it is densely defined and closed. Consequently, Cf is defined and is the zero element for *all* f in $\mathfrak{H}$. Thus, C is the (bounded) zero operator in $\mathfrak{H}$, a contradiction.)

In order to prove the second assertion of (ii), suppose, if possible, that the assertion is false. That is, suppose that B has an eigenvalue z_0, where $\lambda + \mu z_0 \neq 0$, with an associated eigenfunction f in $\mathfrak{D}_A$. Proceeding as in the proof of (i) one obtains

$$z_0 K(z_0) A f = z_0 K(z_0) A f + (\lambda + \mu z_0) f, \tag{24}$$

corresponding to (23). Relation (24) leads to a contradiction and the proof of Theorem III is complete.

THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] B. Fuglede, "A commutativity theorem for normal operators," *Proceedings of the National Academy of Sciences*, vol. 36 (1950), pp. 35-40.

[2] P. R. Halmos, "Commutativity and spectral properties of normal operators," *Acta Scientiarum Mathematicarum*, vol. 12 (1950), pp. 153-156.

[3] B. v. Sz. Nagy, *Spektraldarstellung linearer transformationen des Hilbertschen Raumes*, Ann Arbor, 1947.

[4] C. R. Putnam, "On commutators of bounded matrices," *American Journal of Mathematics*, vol. 73 (1951), pp. 127-131.

[5] F. Rellich, "Der Eindeutigkeitssatz für die Lösungen der quantenmechanische Vertauschungsrelationen," *Göttinger Nachrichten*, 1946, pp. 107-115.

[6] H. Wielandt, "Ueber die Unbeschränktheit der Operatoren der Quantenmechanik," *Mathematische Annalen*, vol. 121 (1949), p. 21.

[7] A. Wintner, "On non-singular bounded matrices," *American Journal of Mathematics*, vol. 54 (1932), pp. 145-149.

[8] ———, "On normal inertia," *Studia Mathematica*, vol. 8 (1939), pp. 135-140.

[9] ———, "The unboundedness of quantum-mechanical matrices," *Physical Review*, vol. 71 (1947), pp. 738-739.

SOME RETRACTION PROPERTIES OF THE ORBIT DECOMPOSITION SPACES OF PERIODIC MAPS.*

By E. E. Floyd.

Let X be a finite-dimensional locally compact metric space. Let T be a periodic map on X. There is generated a decomposition of X whose elements are the orbits $[T^i(x) \mid i = 0, 1, \cdots]$ of points x in X. There is then generated a corresponding decomposition space X^* called the modular space or the orbit decomposition space of the pair X, T. We show that several connectedness and local connectedness properties are transmitted from X to X^*. As a consequence, we prove that if X is either an absolute neighborhood retract or an absolute retract, then X^* has the same property.

Our results depend in a very fundamental fashion on the theory of P. A. Smith concerning periodic maps [6, 7, 8, 9]. Hence we assume on the part of the reader a general knowledge of his methods and results.

Unless otherwise stated, *we assume that X is a locally compact n-dimensional metric space, $n < \infty$, and that T is a periodic map of X onto X of prime period p.* It will be pointed out in Theorem 5 that the preceding results hold when p is not prime. Denote by X^* the collection of all orbits in X, and by $\Lambda : X \rightarrow X^*$ the function which assigns to each x in X the orbit containing x. If we give to X^* the customary decomposition topology, then Λ is a continuous open map of X onto X^*. It may be shown that the map Λ is also compact; i. e., if $C \subset X^*$ is compact, then $\Lambda^{-1}(C)$ is compact. We denote by L the fixed point set of T, and define $L^* = \Lambda(L)$.

If X is a locally compact space, and G an abelian group, then X is *acyclic over G* if every Čech cycle on a compact subset of X over G bounds on a compact subset of X (we shall agree that each coordinate of a 0-cycle has coefficient sum 0). We denote by I the group of integers, and by I_p the group of integers modulo p.

THEOREM 1. *If X is acyclic over I_p, then so is X^*.*

Proof. We shall base the homology theory of X on the complete family of special coverings of X [8, p. 136]. All projections will be invariant projections [8, p. 138]. The homology theory for X^* will be based on the coverings of the type $\mathfrak{U}^* = \Lambda\mathfrak{U}$, where $\mathfrak{U}$ is special.

---

* Received October 9, 1950.

Let W^* be a compact set in X^*. We construct compact sets V^* and U^* with $W^* \subset V^* \subset U^*$, with every cycle in $V^* \cap L^*$ bounding in $U^* \cap L^*$, and with every cycle in W^* bounding in V^* mod $V^* \cap L^*$. As we point out at the end of the proof, this is sufficient to prove that every cycle in W^* bounds in U^*.

Let $W = \Lambda^{-1}(W^*)$. We pick V^* so that every ρ-cycle in W ρ-bounds in $V = \Lambda^{-1}(V^*)$ (for definition of a ρ-cycle, see [7, pp. 361-362]). That this is possible may be seen by examining the proof of a theorem of Smith [6, Theorem II]. Then we pick U^* so that every cycle in $V \cap L$ bounds in $U \cap L$, where $U = \Lambda^{-1}(U^*)$. This is possible since L is acyclic over I_p [6, Theorem II].

The sets U^*, V^*, W^* have the properties of the second paragraph of this proof. It is clear that every cycle in $V^* \cap L^*$ bounds in $U^* \cap L^*$, since Λ is a homomorphism on L. Now let $\gamma = \{\gamma(\mathfrak{U}^*)\}$ be a cycle in W^*. Consider the cycle $\Gamma = \{\Gamma(\mathfrak{U})\} = \{\Lambda^{-1}\gamma(\mathfrak{U}^*)\}$ [8, p. 140]. Following Smith, we let $\delta = 1 - T$ and $\sigma = 1 + T + \cdots + T^{p-1}$. Then Γ is a δ-cycle in W, and hence Γ δ-bounds in V. Moreover, there exist chains $X(\mathfrak{U})$ in W with $\Lambda X(\mathfrak{U}) = \gamma(\mathfrak{U})$ and with $\Gamma(\mathfrak{U}) = \sigma X(\mathfrak{U})$. Since Γ δ-bounds in V, it follows from a theorem of Smith [8, p. 144] that $\gamma \sim 0$ in V^* mod $V^* \cap L^*$. Hence the conditions of the second paragraph are fulfilled.

To prove that every cycle in W^* bounds in U^*, note the following diagram where all homomorphisms are inclusion homomorphisms, and I_p is the coefficient group:

$$\begin{tikzcd}
 & & H_n(W^*) \arrow[dl, "j"'] \\
H_n(V^* \cap L^*) \arrow[r, "k"] \arrow[d, "l"] & H_n(V^*) \arrow[r, "i"] \arrow[d, "l_1"] & H_n(V^*, V^* \cap L^*) \\
H_n(U^* \cap L^*) \arrow[r, "k_1"] & H_n(U^*) \arrow[r] & H_n(U^*, U^* \cap L^*)
\end{tikzcd}$$

Let $x \varepsilon H_n(W^*)$. Then, by construction, $ij(x) = 0$. By exactness, there exists $y \varepsilon H_n(V^* \cap L^*)$ with $k(y) = j(x)$. But $l(y) = 0$. Hence $l_1 k(y) = k_1 l(y) = 0$. Then $l_1 j(x) = 0$. We have proved that every cycle in W^* bounds in U^*, and the theorem follows.

As is customary, we say that a locally compact space M is *lc* over the group G if given a neighborhood U of $x \varepsilon M$, there exists a neighborhood V of x with $V \subset U$ and with every cycle in V bounding in U.

THEOREM 1′. *If X is lc over I_p, so also is X^*.*

Proof. We indicate the minor changes we make in the proof of the preceding theorem. We need only prove that X^* is *lc* at points $y \varepsilon L^*$ since

Λ is a local homomorphism at points of $X - L$. Using the notation of the previous theorem, we have U^* preassigned. Since L is *lc* [9, p. 703], we may pick V^* so that every cycle in $V \cap L$ bounds in $U \cap L$. We then pick W^* so that every ρ-cycle in W ρ-bounds in V. The proof then goes exactly as before.

THEOREM 2. *If X is acyclic over I and over I_p, then so also is X^*.*

Proof. Let W^* be a compact set in X^*. Let $W = \Lambda^{-1}(W^*)$. We pick a compact set V^* with $V^* \supset W^*$ and with every cycle in W^* over I_p bounding in V^*. We may suppose, moreover, that every cycle in W over I bounds in $V = \Lambda^{-1}(V^*)$. Let U^* be a compact set with $U^* \supset V^*$ and with every cycle in V^* over I_p bounding in U^*. We shall prove that every cycle in W^* over I bounds in U^*.

We first note that if z is a cycle over I in W^*, then $pz \sim 0$ in V^*. For $\Lambda^{-1}z$ is a cycle in W with $\Lambda\Lambda^{-1}z = pz$ [8, p. 141]. Since $\Lambda^{-1}z$ bounds in V, then pz bounds in V^*.

If z is a chain over the integers, we denote by $r(z)$ the chain over I_p obtained by replacing each coefficient of z by the corresponding integer mod p. Let K be a complex and z a cycle in K over I. Suppose c is a chain with $\phi c = pz$, where ϕ is the boundary operator. Then $r(c)$ is a cycle over I_p. Moreover if $z \not\sim 0$ then $r(c) \not\sim 0$ [1, p. 220 ff.].

Since every cycle in V^* over I_p bounds in U^*, then if $\mathfrak{U}^*$ is a covering of X^* there exists a covering $\mathfrak{V}^*$ such that if π denotes a projection of $\mathfrak{V}$ into $\mathfrak{U}$ then for any cycle $z(\mathfrak{V}^*)$ over I_p in V^* we have $\pi z(\mathfrak{V}^*) \sim 0$ in U^*. Suppose now that $z = \{z(\mathfrak{U}^*)\}$ is a cycle over I in W^*. We have $pz \sim 0$ in V^*. Let $c(\mathfrak{V}^*)$ be a chain in V^* with $\phi c(\mathfrak{V}^*) = pz(\mathfrak{V}^*)$. Then $\phi \pi c(\mathfrak{V}^*) = p\pi z(\mathfrak{V}^*)$. But $r[\pi c(\mathfrak{V}^*)] = \pi r[c(\mathfrak{V}^*)] \sim 0$ in U^*. Hence $\pi z(\mathfrak{V}^*) \sim 0$ in U^* so that $z(\mathfrak{U}^*) \sim 0$ in U^*. Then $z \sim 0$ in U^* so that the theorem follows.

THEOREM 2′. *If X is lc over I, then so also is X^*.*

Proof. We note that X is also *lc* over I_p [2, p. 11], and hence so also is X^*. The proof is then practically the same as that of Theorem 2, where U^* is now a compact closed neighborhood of a point of L^* and V^* and W^* are then picked so that U^*, V^*, W^* have the properties of the above proof.

THEOREM 3. *Suppose X is arcwise connected, and that $L \neq 0$. Then if X is simply connected, so also is X^*.*

Proof. Let $x \varepsilon L$ and let $y = \Lambda(x)$. Let $f: I \rightarrow X^*$ be a path in X^*, where I is the unit interval, such that $f(0) = f(1) = y$. Then we consider Λ as being a light interior map of the compact space $\Lambda^{-1}f(I)$ onto $f(I)$. Then there exists a map $g: I \rightarrow X$ with $\Lambda g = f$ [3]. Then $g(0) = g(1) = x$, since $\Lambda^{-1}(y) = x$. Then g is a closed path in X and hence $g \backsim 0$ in X. Then $f \backsim 0$ in X^* and the assertion follows.

THEOREM 3′. *If X is locally connected and locally simply connected, then so is X^*.*

Proof. The proof is similar to the above, and may be omitted.

THEOREM 4. *If X is an absolute neighborhood retract [an absolute retract], then X^* is also an absolute neighborhood retract [absolute retract.]*

Proof. Suppose X is an *ANR*. Then X is LC^1 and *lc* over I. By Theorems 2 and 3, X^* is LC^1 and *lc* over I. Then by a theorem of Hurewicz [5] we have that X^* is LC^n and hence an *ANR*. Suppose X is an *AR*. Then X is simply connected and acyclic over I and I_p. By Theorem 2, X^* is acyclic over I. Moreover $L \neq 0$ [6]. Then by Theorem 3, X^* is simply connected. It follows, then, that X^* is an *AR* [4].

We shall now point out that we may, in our results, replace the cyclic transformation group of prime order p generated by T by any finite solvable transformation group. Let G be a finite transformation group on X. Let H be a normal subgroup of G. Let X^*_1 be the orbit decomposition space of the pair X, H and $\Lambda_1: X \rightarrow X^*_1$ the corresponding decomposition map. The points of X^*_1 are sets of the form $H(x)$, $x \varepsilon X$. If $G_1 = G/H$, if $h = \{g\}$ is an equivalence class in G_1 and if $H(x) \varepsilon X^*_1$, define $h[H(x)] = (gH)(x) = H(gx) \varepsilon X^*_1$. In this way G_1 acts as a transformation group on X^*_1. Call the resulting orbit decomposition space X^*_2 and denote by $\Lambda_2: X^*_1 \rightarrow X^*_2$ the corresponding decomposition map. Finally, let X^* be the orbit decomposition space of the pair X, G and $\Lambda: X \rightarrow X^*$ the decomposition map.

LEMMA 1. *There exists a homeomorphism S of X^* onto X^*_2 with $S\Lambda = \Lambda_2\Lambda_1$.*

Proof. Let $x \varepsilon X$. Then $\Lambda^{-1}\Lambda(x) = G(x)$. It may be seen, letting $\Gamma = \Lambda_2\Lambda_1$, that $\Gamma^{-1}\Gamma(x) = G(x)$. It follows that the function $\Gamma\Lambda^{-1}$ is a 1-1 function of X^* onto X^*_2. Moreover it is bicontinuous since both Γ and Λ are continuous and open. Setting $S = \Gamma\Lambda^{-1}$, we obtain the conclusion.

THEOREM 5. *Let G be a finite solvable transformation group on X. Let X^* denote the orbit decomposition space of the pair X, G. Then if X is an absolute neighborhood retract [absolute retract], so also is X^*.*

Proof. Let $\Lambda: X \rightarrow X^*$ be the decomposition map. According to the previous lemma, we may write $\Lambda = \Lambda_m \cdots \Lambda_2\Lambda_1$ where each Λ_i is equivalent to the decomposition map generated by a prime periodic map. The theorem then follows from Theorem 4.

Let us also note the following problems:

PROBLEM 1. *Does Theorem 5 hold when G is not solvable?*

PROBLEM 2. *If X is a manifold, does X^* contain about each point arbitrarily small contractible neighborhoods? Does, as a matter of fact, X contain arbitrarily small contractible invariant neighborhoods about each of its fixed points?*

PROBLEM 3. *If X is a 3-manifold, is there an arbitrarily small invariant 2-sphere about each fixed point?*

UNIVERSITY OF VIRGINIA.

BIBLIOGRAPHY.

1. P. Alexandroff and H. Hopf, *Topologie*, Berlin (1935).
2. E. Čech, "Sur la connexité locale d'ordre supérieur," *Compositio Mathematica*, vol. 2 (1935), pp. 1-25.
3. E. E. Floyd, "Some characterizations of interior maps," *Annals of Mathematics*, vol. 51 (1950), pp. 571-575.
4. W. Hurewicz, "Beiträge zur Topologie der Deformationen. II. Homotopie- und Homologiegruppen," *Nederlandsche Akademie van Wetenschappen, Proceedings*, vol. 38 (1935), pp. 521-528.
5. W. Hurewicz, "Homotopie, Homologie und lokaler Zusammenhang," *Fundamenta Mathematicae*, vol. 25 (1935), pp. 467-485.
6. P. A. Smith, "Fixed point theorems for periodic transformations," *American Journal of Mathematics*, vol. 63 (1941), pp. 1-8.
7. ———, "Fixed points of periodic transformations," Appendix B in Lefschetz's *Algebraic Topology*, New York (1942).
8. ———, "Transformations of finite period," *Annals of Mathematics*, vol. 39 (1938), pp. 127-164.
9. ———, "Transformations of finite period. II," *Annals of Mathematics*, vol. 40 (1939), pp. 690-711.

ON THE NON-EXISTENCE OF CONJUGATE POINTS.*

By Aurel Wintner.

1. Let $f(t)$ be a real-valued, continuous function on a t-interval, I, which need not be closed or bounded. Consider only those solutions $x(t)$ of the differential equation

$$x'' + f(t)x = 0 \tag{1}$$

on I which are real-valued and are distinct from the trivial solution ($\equiv 0$). If no solution of (1) has more than one zero on I, then (1) will be called disconjugate on I. This terminiology is suggested by the fact that, in calculus of variations, the disconjugate nature of (1) on I defines the absence of conjugate points on I, if (1) is Jacobi's equation.

The remarks collected in this note will center about the following criterion: (1) is disconjugate on I if and only if there exists on I *some* function $y = y(t)$ possessing a continuous first derivative which satisfies the inequality

$$y'(t) + y^2(t) \leqq -f(t) \tag{2}$$

at every t contained in I.

If the $\leqq$ is refined to an $<$ in (2), the sufficiency of the resulting criterion is contained in a more general result of Bôcher [1], who has applied it in order to derive Sturm's "separation" and "comparison" theorems. But the classical proof of these theorems on (1) is simple enough and, if they are granted, both the necessity and the sufficiency of the criterion (2) follow as immediate corollaries. For the sake of completeness, this will be shown, in a generalized form, in Section **10**.

Sturm's comparison theorem follows by observing that if $f_1(t) \geqq f_2(t)$, and if (2) is satisfied by a y for $f = f_1$, then it is satisfied, by the same y, for $f = f_2$. But it does not seem to have been noticed in the literature that the criterion (2) easily leads to non-trivial results also. For instance, it implies, without the *ad hoc* considerations of the known proofs, and almost as trivialities, the explicit sufficient criteria which go back to Liapounoff (cf. [2]) and Kneser [4], respectively. In addition, other sufficient criteria also follow from the same source.

---

* Received August 20, 1950.

2. First, it will be shown that the two explicit criteria, just mentioned, are manifestations of one and the same situation, the t-range, I, being (α) a bounded interval in the first case and (β) a half-line in the second. If a^+ denotes $\max(0, a)$, these two criteria can be formulated as follows:

The differential equation (1) must be disconjugate on $I = [0, 1]$ if

$$(\alpha) \qquad \int_0^1 f^+(s)\,ds \leqq 4,$$

and it must be disconjugate on $I = (0, \infty)$ if

$$(\beta) \qquad \int_t^\infty f^+(s)\,ds \leqq 1/(4t) \text{ for } 0 < t < \infty$$

(the convergence of the last integral is part of the assumption, since $f^+ \geqq 0$).

Ad (α). Since $f^+(t) \geqq f(t)$, the differential equation (1) is a Sturmian minorant of the differential equation $x'' + f^+(t)x = 0$. On the other hand, $f^+(t) \geqq 0$. Hence, it is sufficient to show that if (α) is replaced by the conditions

$$(3) \qquad \int_0^1 f(s)\,ds \leqq 4,\ f(t) \geqq 0, \text{ where } 0 \leqq t \leqq 1,$$

then (1) is disconjugate on $I = [0, 1]$. To this end, define for $0 < t < 1$ a function $y(t)$ by placing

$$(4) \qquad y(t) = \int_t^1 f(s)\,ds + \begin{cases} t^{-1} - 4 & \text{if } 0 < t \leqq \frac{1}{2}, \\ (t-1)^{-1} & \text{if } \frac{1}{2} \leqq t < 1. \end{cases}$$

This $y(t)$ has a continuous derivative for $0 < t < 1$. If (3) is assumed, it is readily verified that the function (4) satisfies (2) for $0 < t < 1$. Hence, (3) is sufficient in order that (1) be disconjugate on $I = (0, 1)$. Finally, the transition from this open interval I to the corresponding closed interval, $I = [0, 1]$, can be effected by using the continuity of $f(t)$ on $[0, 1]$ and observing that in (2) the sign of inequality must hold on some subinterval of $(0, 1)$ by virtue of (3) and (4).

Ad (β). Corresponding to the transition from (α) to (3), it can be assumed that (β) is replaced by

$$(4') \qquad \int_t^\infty f(s)\,ds \leqq 1/(4t) \text{ and } f(t) \geqq 0, \text{ where } 0 < t < \infty.$$

Then a direct substitution shows that (2) is satisfied by

$$y(t)=\int_t^\infty f(s)\,ds+1/(4t). \tag{5}$$

It is clear from these verifications that both (3) and (4′) can be improved by suitable "logarithmic" functions. In addition, (β) can be modified as follows:

(β^*) If $\int^\infty f(t)\,dt$ is convergent (possible just conditionally) and, as a function of its lower limit of integration, satisfies the inequality

$$\Big|\int_t^\infty f(s)\,ds\Big|\leqq 1/(4t) \tag{4*}$$

on the half-line I, then (1) is disconjugate on I.

In other words, the restriction $f(t)\geqq 0$ can be omitted in (4′). In fact, it is readily seen that the $y(t)$ defined by (5) satisfies (2) by virtue of (4*) alone.

Actually, (4*) can be improved to

$$-\tfrac{3}{4}\leqq t\int_t^\infty f(s)\,ds\leqq\tfrac{1}{4}. \tag{4**}$$

In fact, (2) holds for the function (5) whenever

$$\Big|\int_t^\infty f(s)\,ds+1/(4t)\Big|\leqq 1/(4t^2)^{\frac{1}{2}}.$$

But this inequality can be written in the form

$$-1/(2t)\leqq\int_t^\infty f(s)\,ds+1/(4t)\leqq 1/(2t),$$

which is equivalent to (4**).

It should be mentioned that the constant $-\frac{3}{4}$ occurring in (4**) will arise in Section **9** in a different context.

3. For $I=(0,\infty)$, a new sufficient condition, of a type quite different from (4), is as follows:

$$(6) \qquad [\int_t^\infty f(s)\,ds]^2 \leqq f(t)/4, \text{ where } 0 < t < \infty.$$

The sufficiency of (6) is readily verified by choosing $y(t) = 2 \int_t^\infty f(s)\,ds$ in (2).

The constant factor $\frac{1}{4}$ occurring in (4′), (4*), (6) is the best absolute constant. In fact, if $f(t) = c/t^2$, then (4′), (4*), (6) are satisfied when $4c \leqq 1$, but the $\frac{1}{4}$ must be replaced by a somewhat larger constant if $4c = 1 + \epsilon \rightarrow 1 + 0$. Hence it is sufficient to ascertain that the case $f(t) = c/t^2$ of (1) cannot be disconjugate on $(0, \infty)$ if $4c > 1$. But this is well-known (Kneser). In fact, the case $f(t) = c/t^2$ of (1) is satisfied, for any real c and for $0 < t < \infty$, by $x(t) = t^b$, and therefore by the real part of t^b as well, if $b(b-1) + c = 0$. Since this quadratic equation has real roots only if $4c \leqq 1$, it follows that the real solutions $x(t)$ must have zeros clustering at $t = \infty$ if $4c > 1$.

If the half-line $I = (0, \infty)$ is replaced by a bounded interval, say by $I = [0, a]$, then a sufficient condition corresponding to (6) is

$$(7) \qquad [\int_0^t f(s)\,ds]^2 \leqq f(t)/4, \text{ where } 0 \leqq t \leqq a.$$

In fact, (7) implies that (2) is satisfied, for $0 \leqq t \leqq a$, by $y(t) = -2 \int_0^t f(s)ds$.

4. Let I be a half-line reaching to ∞, say $I = (t_0, \infty)$. If (1) has a solution the zeros of which cluster at infinity, then (1) is called oscillatory, and otherwise non-oscillatory. It is clear from Sturm's separation theorem that if (1) is oscillatory, then every solution of (1) has zeros clustering at infinity.

For the same reason, it is easy to see that if (1) is non-oscillatory, then there must exist a sufficiently large t^0 ($\geqq t_0$) having the property that no solution of (1) possesses more than one zero on the half-line (t^0, ∞). Consequently, (1) is non-oscillatory if and only if there exists on *some* half-line *some* function $y(t)$ possessing a continuous first derivative which satisfies (2) (on that half-line).

For instance, it follows from (β*) in Section 2 that if $\int^\infty f(s)\,ds$ is convergent (possibly just conditionally) and, when considered as a function of its lower limit, satisfies the condition

$$\limsup_{t\to\infty} t\left|\int_t^\infty f(s)\,ds\right| < \tfrac{1}{4}, \tag{8}$$

then (1) is non-oscillatory. In particular, (1) must be non-oscillatory if

$$\int_t^\infty f(s)\,ds = o(t^{-1}) \text{ as } t\to\infty, \tag{9}$$

where the (possible just conditional) convergence of the integral is part of the assumption.

A criterion of a different type is as follows: If the $f(t)$ in (1) can be written as the sum of two functions, $f_1(t)$ and $f_2(t)$, which are such that

$$\limsup_{t\to\infty} f_1(t) < 0 \text{ and } \int^\infty f_2(t)\,dt \text{ converges}$$

(possibly just conditionally), then (1) is non-oscillatory.

In fact, $y(t) = \int_t^\infty f_2(s)\,ds$ will satisfy (2) for large t. For, since $f = f_1 + f_2$, this $y(t)$ reduces (2) to

$$\left[\int_t^\infty f_2(s)\,ds\right]^2 \leqq -f_1(t),$$

a condition which is satisfied for large t, since the integral on the left tends to 0, while $-f_1(t)$ on the right has a positive lower bound. Needless to say, this sufficient condition for non-oscillatory behavior cannot be obtained from Sturm's comparison theorem. In fact, $\limsup f_1 < 0$ and the convergence of the integral of f_2 (which, incidentally, can readily be lightened) do not preclude the possibility $\limsup f \geqq 0$ (or, for that matter, $\limsup f = \infty$) for $f = f_1 + f_2$.

5. With an arbitrary choice of the lower limit of integration, put

$$F(t) = \int^t f(s)\,ds \tag{10}$$

and suppose that

$$F(t) \to \infty \text{ as } t \to \infty \tag{11}$$

(but $f \geqq 0$ is *not* assumed). Then (1) must be oscillatory. Actually, the same is true if (11) is relaxed to

$$(12)\qquad \int^{t} F(s)\,ds/t \to \infty \text{ as } t \to \infty$$

(cf. [5]), but the proof in the case (11) will suffice as an illustration of the use of (2) in the direction which is the opposite of that followed in Sections **2-3**.

Suppose that the assertion is false. Then, although (11) is satisfied, (1) is non-oscillatory, that is, that there exists a $y(t)$ for large t. For this $y(t)$, a quadrature of (2), when combined with (10) and (11), shows that

$$(13)\qquad y(t) + \int^{t} y^2(s)\,ds \to -\infty \text{ as } t \to \infty.$$

But (13) implies that $y(t) \to -\infty$, hence $|\, y(t) \,| \to \infty$. Consequently, the integral in (13) tends to ∞ much stronger than $|\, y(t) \,|$. Accordingly, (13) contains a contradiction.

This proves that (1) must be oscillatory if (11) holds. On the other hand, (1) need not be non-oscillatory if (11) is replaced by

$$(14)\qquad F(t) \to -\infty \text{ as } t \to \infty.$$

In order to see this, suppose that $f(t)$ is periodic, and let λ denote the constant term in the Fourier series of $f(t)$. Then it is clear from (10) that (11) or (14) is satisfied according as $\lambda > 0$, $\lambda < 0$. But it is not true that $\lambda = 0$ is the cut between the oscillatory and non-oscillatory cases of (1). In fact, even Mathieu's

$$(15)\qquad x'' + (\lambda + a^2 \cos t)x = 0, \text{ where } a \neq 0,$$

is known to be oscillatory for $\lambda > -\lambda^*$, where $\lambda^* = \lambda^*(a)$ is positive (cf. (23) below).

Instead of the periodicity of $f(t)$, suppose only that $f(t)$ has, for large t, a mean-value, $M(f)$. According to (10), this means that there exists a number, $M(f)$, for which

$$(16)\qquad F(t)/t \to M(f) \text{ as } t \to \infty.$$

If $M(f) > 0$, then (12) is satisfied, hence (1) is oscillatory. If $M(f) < 0$, then (1) may or may not be non-oscillatory; cf. (14) and (15). But if $M(f) < 0$, and if the convergence in (16) is so rapid that the absolute value of the error does not exceed $[-M(f)]^{\frac{1}{2}}/t$ when t is large, then (1) must be non-oscillatory.

In fact, the assumption is that, if the lower limit of integration is suitably chosen, then, from a certain t onward,

$$(17) \qquad \left| \int^{t} f(s)\,ds - M(f)t \right| \leqq [-M(f)]^{\frac{1}{2}};$$

cf. (10) and (16). But a direct substitution shows that (2) is satisfied by

$$(18) \qquad y(t) = -\int^{t} f(s)\,ds + M(f)t$$

for large t, if (17) is assumed.

6. This proof also shows that if $f(t)$ is defined only on the interval $0 \leqq t \leqq 1$, and satisfies the inequality

$$(19) \quad |F(t) - F(1)t| \leqq [-F(1)]^{\frac{1}{2}}, \text{ where } F(t) = \int_{0}^{t} f(s)\,ds,\ 0 \leqq t \leqq 1,$$

then (1) is disconjugate on $I = [0, 1]$. In fact, the assumption (19), which corresponds to (17), assures that (2) is satisfied, for $0 \leqq t \leqq 1$, by the $y(t)$ which results if the lower limit of integration is chosen to be $s = 0$ in (18).

Clearly, (19) can be written in the form

$$(20) \quad \left| \int_{0}^{t} \{f(s) - \lambda\}\,ds \right| \leqq (-\lambda)^{\frac{1}{2}}, \qquad 0 \leqq t \leqq 1,\ \lambda = \int_{0}^{1} f(t)\,dt.$$

On the other hand, an application of Schwarz's inequality shows that (20) is surely satisfied if

$$(21) \qquad \int_{0}^{1} [f(s) - \int_{0}^{1} f(t)\,dt]^2 ds \leqq -\int_{0}^{1} f(t)\,dt.$$

Finally, (21) can be interpreted by saying that the mean-value of $f(t)$ on $[0, 1]$ is non-positive and majorizes the squared standard derivation of $f(t)$. Hence, such a majorization of the standard deviation of $f(t)$ on $[0, 1]$ is sufficient to make (1) disconjugate on $[0, 1]$.

A corollary is that if $f(t)$ is periodic, of period 1, and satisfies (21), or the more general condition (20), then (1) must be non-oscillatory.

Needless to say, (20) requires a preponderance of $c_0 = -|c_0|$ in the Fourier series $f(t) \sim \Sigma\, c_n \exp(2\pi int)$, where $n = 0, \pm 1, \cdots$ and $c_{-n} = \overline{c_n}$. In particular, (21) is equivalent to $|c_1|^2 + |c_2|^2 + \cdots \leqq -\frac{1}{2}c_0$, by Parseval's relation.

Accordingly, if $\phi(t)$ is a continuous, periodic function of period 1 and of mean-value 0, and if λ^* denotes the cut between those λ-values for which the differential equation

$$x'' + \{\lambda + \phi(t)\}x = 0, \text{ where } \phi(t) \sim \sum_{n=-\infty}^{\infty}{}' \gamma_n \exp(2\pi\, int),\ n \neq 0, \tag{22}$$

is oscillatory and those for which it is non-oscillatory, then

$$\lambda^* \geqq -2\sum_{n=1}^{\infty} |\gamma_n|^2 \tag{23}$$

(and (22) is non-oscillatory also for the λ-value which is the expression on the right-hand side of (23)).

An upper limitation of λ^*, corresponding to the lower limitation (23), does not follow. All that is clear in this direction is the inequality $\lambda^* \leqq 0$, which is contained in the criterion (11).

7. In order to avoid an interruption of the proofs below, it will be convenient to isolate the following remark:

For large positive t, let $f(t)$ be defined in such a way that the improper integral

$$\int^{\infty} f(t)\,dt \text{ is convergent} \tag{24}$$

(possible just conditionally), and that (1) is non-oscillatory. Then

$$\int^{\infty} l^2(t)\,dt < \infty \tag{25}$$

must hold for the logarithmic derivative, $l = x'/x$, of every solution, $x = x(t)$, of (1). In addition,

$$l(t) \rightarrow 0 \text{ as } t \rightarrow \infty. \tag{26}$$

Finally,

$$l(t) = \int_t^{\infty} \{l^2(s) + f(s)\}\,ds. \tag{27}$$

It is understood that the lower limits of integration in (25) and (27) are supposed to be chosen large enough, larger than the last zero of $x(t)$. Since (1) is supposed to be non-oscillatory, there exists such a zero for every $x(t)$ (if there is at all a zero). Let it be denoted by t_0 (if it exists; otherwise choose t_0 arbitrarily), and let $t_0 < t < \infty$. Then $l(t) = x'(t)/x(t)$ can be formed and, as seen if (1) is divided by x, satisfies the differential equation

$$l'(t) + l^2(t) + f(t) = 0 \tag{28}$$

(Riccati). But if a quadrature is applied to (28), it follows from (24) that

$$l(T) + \int^{T} l^2(t)\,dt$$

must tend, as $T \to \infty$, to a finite limit. Hence, a repetition of the argument used after (13) shows that the negation of the truth of (25) leads to a contradiction. Consequently, (25) implies that $l(t)$ must tend, as $t \to \infty$, to a limit. Hence, (26) follows from (25). Finally, (27) follows from (26) and (28).

In the proof of (25)-(28), it was not assumed that

$$f(t) \geqq 0. \tag{29}$$

If (29) is assumed, then the hypothesis (24) can be omitted by virtue of the other hypothesis, that requiring that (1) be non-oscillatory. In other words, if (29) is assumed but (24) does not hold, then (1) must be oscillatory. In fact, (29) and the negation of (24) imply that (10) satisfies (11).

8. Without assuming (29), the following criterion will now be proved:

If (24) is assumed, and if the integral (24), when considered a function of its lower limit, satisfies the condition

$$\int^{\infty} \exp\{-2 \int^{t} [\int_{u}^{\infty} f(v)\,dv]\,du\}\,dt < \infty, \tag{30}$$

then (1) must be oscillatory (the choice of the unspecified lower limits of integration in (30) is immaterial).

This criterion for the oscillatory behavior of (1) is quite different from the known criteria, since *individual* t-values do not occur in (30). In other words, $f(t)$ and its repeated indefinite integrals can be of arbitrarily "irregular growth" (even if $f(t)$ is non-negative, which is not assumed). Correponding to the fact that t occurs in (30) only beneath an integral sign, it would be inconvenient to deduce (30) from the general criterion (2). But the latter might lead to a final result, whereas it will remain undecided whether the 2 in (30) is the best absolute constant. Sure is only that the 2 cannot be more than doubled. In fact, if $f(t) = c/t^2$, then (30) is satisfied when $c > \frac{1}{2}$, whereas the case $f(t) = c/t^2$ of (1) is oscillatory if and only if $c > \frac{1}{4}$ (cf. Section **3**).

In order to prove that (1) is oscillatory under the assumptions (24) and (30), suppose that (1) is non-oscillatory. Then, if $x(t)$ is any solution, (27) is applicable for large t. In particular, since $l = x'/x$,

$$d \log x(t)/dt > \int_t^\infty f(s)\,ds \tag{31}$$

and $x(t) \neq 0$ hold from a certain t onward. In addition, it can be assumed that $x(t) > 0$, since if $x(t)$ is a solution of (1), then $-x(t)$ is. Hence, if a quadrature is applied to (31), it follows from the assumption (30) that

$$\int^\infty [x(t)]^{-2}\,dt < \infty \tag{32}$$

holds for every solution of (1), if the lower limit of integration (34) exceeds the last zero of the solution. But this contains a contradiction, since, as shown by Hartman [3], every (non-oscillatory) differential equation (1) must have some solution violating (32).

It will be noted that this proof seems to use, via (31), just about the half of the content of (27). Actually, this is not quite correct, as seen by evaluating the integrals involved for the case $f(t) = c/t^2$ of an arbitrary $c > \frac{1}{4}$.

9. Along the lines of the preceding proof, it is easy to obtain a sufficient condition for the non-existence of an "eigensolution" (of class L^2), that is, of a solution ($\not\equiv 0$) satisfying

$$\int^\infty x^2(t)\,dt < \infty. \tag{33}$$

Suppose first that $f(t)$ satisfies (29) but (1) is non-oscillatory. Then, since $x(t)$ can be supposed to be positive for large t, it is seen from (1) that, for reasons of convexity, $x(t)$ must have a positive lower bound as $t \to \infty$, and so (33) is impossible.

Next, suppose that (29) is replaced by

$$f(t) \leqq 0. \tag{34}$$

Then (1) must be non-oscillatory (in fact, (34) implies that (2) holds for $y(t) \equiv 0$). But (1) may or may not have a solution satisfying (33), even if (24) is assumed. In fact, if $f(t) = c/t^2$, then the explicit form of the solutions, described in Section **3**, shows that (1) has a solution satisfying (33) if and only if $c < -\frac{3}{4}$.

Without asuming (34), it is easy to see that, if (1) is non-oscillatory and if $f(t)$ satisfies (24) and

$$(35)\qquad \int^{\infty} \exp\{2 \int^{t} [\int_{u}^{\infty} f(v)\,dv]\,du\}\,dt = \infty,$$

then (1) cannot have a solution satisfying (33).

In fact, since (27) is applicable, (31) holds for every solution of (1). But the application of a quadrature on (31) shows that (33) is impossible if (35) is assumed.

For instance, a non-oscillatory differential equation (1) cannot have a solution satisfying (33) if (24) is assumed and

$$(36)\qquad t\int_{t}^{\infty} f(s)\,ds \geqq -\tfrac{1}{2}$$

holds for large t. In fact, (36) implies that the function of t which is integrated in (35) is minorized by a constant multiple of $\exp\{2(-\frac{1}{2}\log t)\} = t^{-1}$; hence (35) is satisfied. In particular, (33) cannot hold for any solution of (1) if $f(t)$ satisfies (34), (24) and (36). It remains undecided whether the $-\frac{1}{2}$ in (36) is the best constant (whether (34) is or is not assumed). The example $f(t) = c/t^2$, mentioned after (34), shows that the $-\frac{1}{2}$ in (36) cannot be improved to any constant which is less than $-\frac{3}{4}$.

The criterion (36) is applicable only under the proviso that (1) be non-oscillatory. This proviso is sure to be satisfied if

$$(37)\qquad t\int_{t}^{\infty} f(s)\,ds \leqq \tfrac{1}{4}.$$

In fact, (36) and (37) imply both of the inequalities (4**) of Section 2, and (4**) was seen to be sufficient in order that (1) be non-oscillatory. Accordingly, (1) cannot have a solution satisfying (33) if $f(t)$ is subject to (36) and (37) (where (24) is part of the assumption).

In particular, this will be the case if $f(t)$ is subject to (4*), since (4*) implies both (36) and (37), Still more particularly, it is sufficient to subject $f(t)$ to

$$(38)\qquad \int_{t}^{\infty} f(s)\,ds = o(t^{-1}) \text{ as } t\to\infty,$$

since this o-condition implies (4*) for large t.

10. On a fixed t-range I, replace (1) by

$$(p(t)x')' + f(t)x = 0, \tag{39}$$

where $p(t)$ is positive and, as $f(t)$, continuous on I. Let N be defined by the property that every solution $x = x(t) \not\equiv 0$ of (39) has at most N, and some solution has actually N, zeros on I. Then $1 \leqq N \leqq \infty$. In fact, $N = 0$ is impossible, since $x(t_0) = 0$ (along with any $x'(t_0) \neq 0$) can be assigned as an initial condition at a point t_0 of I. If $N = 1$, then (39) will be called disconjugate on I.

As a generalization of the criterion (2), where $p(t) \equiv 1$, it is easy to see that (39) is disconjugate on I if and only if there exists on I *some* function $y = y(t)$ possessing a continuous first derivative and satisfying the inequality $R(y) \leqq 0$ at every point of I, where R denotes the Riccati operator associated with (39), that is,

$$R(y) = y' + y^2/p + f. \tag{40}$$

It is well-known that

(i) if $N = 1$, then (39) possesses solutions which have no zeros at all on I (Jacobi), and that

(ii) a function $y = y(t)$ having a continuous first derivative on I is a solution of the non-linear differential equation $R(y) = 0$ if and only if $y(t)/p(t)$ is the logarithmic derivative of some solution $x(t) \not\equiv 0$ of the linear differential equation (39) on I (Riccati).

Since $p(t) > 0$, and since $x(t)$ and $x'(t)$ cannot vanish simultaneously for any solution $x(t) \not\equiv 0$ of (39) on I, it follows that a solution $y(t)$ of $R(y) = 0$ cannot exist on the whole of I unless $N = 1$, and that this necessary condition, $N = 1$, is sufficient as well; cf. (i). But if $N \neq 1$, then any closed t-interval containing two zeros of a solution ($\not\equiv 0$) of (39) must contain at least one zero of any other solution (Sturm). It follows that

(iii) the non-linear differential equation $R(y) = 0$ does or does not possess some solution $y(t)$ which exists on the whole of I according as $N = 1$ or $N > 1$ (this, in turn, implies that (ii) contains (i) as a corollary).

Let (39), $R^*(.)$ and N^* denote what result if $f(t)$ in (39), (40) and the definition of N is replaced by another continuous function, $g(t)$, while $p(t)$ and I are unchanged. According to Sturm, $N \leqq N^*$ holds whenever $f(t) \leqq g(t)$ on I. If this is combined with the identity $R(.) - R^*(.) = f - g$, it follows from (i), (ii) and (iii) that $N = 1$ if and only if there exists on I some function $y = y(t)$ possessing a continuous first derivative and satisfying $R(y) \leqq 0$ on I; q. e. d.

Remark. Let (39_*), $R_*(.)$ and N_* denote what result if $p(t)$ in (39), (40) and the definition of N is replaced by another positive, continuous function, $q(t)$, while $f(t)$ and I are left unchanged. Then $N_* \leqq N$ whenever $p(t) \leqq q(t)$ holds on I. The customary proofs of this theorem of Sturm are more elaborate than is that of his inequality $N \leqq N^*$, used above (the proof of $N \leqq N^*$ now found in textbooks depends on an *ad hoc* identity of Picone).

Actually, it is clear from the definitions of $R(.)$ and $R_*(.)$ that $R(.) \geqq R_*(.)$ whenever $0 < p \leqq q$. In view of the above criterion, this implies that (39_*) must be disconjugate on I if (39) is. This means that $N_* \leqq N$, provided that $N = 1$. But the latter proviso can readily be removed if I is subdivided in a suitable manner.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Bôcher, *Leçons sur les méthodes de Sturm*, Paris, 1917 (Collection Borel), chap. III.

[2] G. Borg, "On a Liapounoff criterion of stability," *American Journal of Mathematics*, vol. 71 (1949), pp. 67-70.

[3] P. Hartman, "On differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[4] A. Kneser, "Untersuchungen über die reellen Nullstellen der Integrale linearer Differentialgleichungen," *Mathematische Annalen*, vol. 42 (1893), pp. 409-435; cf. E. Hille, "Non-oscillation theorems," *Transactions of the American Mathematical Society*, vol. 64 (1948), pp. 234-252.

[5] A. Wintner, "A criterion of oscillatory stability," *Quarterly Journal of Applied Mathematics*, vol. 7 (1949), pp. 115-117.

ON THE CLASSICAL TRANSCENDENTS OF MATHEMATICAL PHYSICS.[*]

By PHILIP HARTMAN and AUREL WINTNER.

Introduction. On an x-interval (a, b), let $r(x)$, $q(x)$ and the first derivative of a positive $p(x)$ be completely monotone, that is, let

$$(1) \qquad (-1)^n d^n y/dx^n \geqq 0 \text{ on } (a, b) \qquad (n = 0, 1, \cdots)$$

for $y = r, q, p'$, where it is understood that (a, b) is open and need not be bounded. It was shown in [1] that the differential equation

$$(2) \qquad p(x)y'' + r(x)y' - q(x)y = 0$$

must have a solution $y(x)$ which is completely monotone on (a, b) (here and in the sequel, the trivial solution, $y(x) \equiv 0$, is not called a solution). In the particular case $(a, b) = (0, \infty)$, it follows from the Hausdorff-Bernstein theorem that if, $r(x)$, $q(x)$ and the first derivative of a function

$$(3) \qquad p(x) > 0$$

satisfy the inequalities (1) on the half-line $x > 0$, then (2) must have a solution $y(x)$ representable in the form

$$(4) \qquad y(x) = \int_0^\infty e^{-xt}\, d\phi(t), \qquad \text{where } 0 \not\equiv d\phi(t) \geqq 0,$$

on the half-line $x > 0$. For the particular case

$$(5) \qquad x'' - q(x)y = 0$$

of (2), cf. [5].

The purpose of this paper is an application of this result to the case of certain classical linear differential equations occurring in mathematical physics. For instance, there will be considered (§ 1) the so-called toroidal functions which, in the main, are certain normalizations of the associated Legendre functions. The latter are not hypergeometric functions. Applications will, however, be given (§ 2) to the hypergeometric differential equation, and Kummer's confluent case will also be considered (§ 3).

---

[*] Received June 27, 1950.

What results in these cases for the range of an x-half-line is an integral representation of the form (4), where x is a certain elementary function of the independent variable in the standard forms of the differential equations. The results are mere existence statements, since we cannot determine the explicit form of the respective functions $\phi(t)$ in (4). But the range of the hypergeometric or confluent parameters which are treated in this manner are not those occurring in the classical integral representation of Euler and Kummer.

When the x-range is an interval contained between the two finite singular points of the hypergeometric equation, there will result (§ 2) statements on the complete monotony of certain solutions on part of that interval.

It may be mentioned that if $p > 0$, and if $y = r, q, p'$ satisfy (1) for $0 < x < \infty$, then, in general, a solution of (2), representable in the form (4), is unique up to a constant factor. For, if (2) has two linearly independent solutions of the form (4), then all solutions of (2) tend to finite limits as $x \rightarrow +\infty$. For example, if $p(x) \equiv 1$ and $q(x) > 0$ for $x \geqq X$, it follows from a criterion of Weyl [3], p. 40, that a necessary and sufficient condition in order that all solutions of (2) tend to limits is

$$\int_X^\infty \exp(-\int_X^s r(\xi)\,d\xi) \int_X^s q(t) \exp(\int_X^t r(\eta)\,d\eta)\,dt\,ds < \infty. \tag{6}$$

Thus (6) is a necessary condition in order that all solution of (2) be representable in the form (4), when $p \equiv 1$ and $q > 0$. In particular, if

$$q(x) \geqq \text{const.} > 0 \text{ for large } x, \tag{7}$$

then (6) cannot hold, and not all solutions of (2) (or their negatives) are representable in the form (4). For the integral (6) is minorized by a constant multiple of

$$\int_{X+1}^\infty \exp(-\int_X^s r(\xi)\,d\xi) \int_{s-1}^s \exp(\int_X^t r(\eta)\,d\eta)\,dt\,ds$$
$$= \int_{X+1}^\infty \int_{s-1}^s \exp(-\int_t^s r(\xi)\,d\xi)\,dt\,ds.$$

Since $s - 1 \leqq t \leqq s$ and since $r(x)$ is bounded as $x \rightarrow +\infty$, it follows that (6) cannot hold.

1. Toroidal functions. If in the differential equation of the associated Legendre polynomials

$$(8)\qquad (z^2-1)d^2y/dz^2+2zdy/dz-[\nu(\nu+1)+\mu^2/(z^2-1)]y=0$$

the substitution

$$(9)\qquad \nu=n-\tfrac{1}{2},\qquad \mu=m,\qquad z=\cosh x$$

is made, then there results the differential equation

$$(10)\qquad d^2y/dx^2+(\cosh x/\sinh x)dy/dx-[n^2-\tfrac{1}{4}+m^2/\sinh^2 x]y=0$$

for the toroidal functions; cf., e. g., [2], p. 75.

(I) *If* $n^2-\frac{1}{4}\geqq 0$, *then there exists a toroidal function* (*a solution of* (10)) *representable in the form* (4) for $0<x<\infty$. *Furthermore, this toroidal function is unique, up to a constant factor, if and only if* $n^2-\frac{1}{4}>0$.

Proof of (I). A comparison of (10) and (2) shows that $p(x)\equiv 1$,

$$(11)\qquad r(x)=\cosh x/\sinh x \text{ and } q(x)=n^2-\tfrac{1}{4}+m^2/\sinh^2 x.$$

Thus, $p'(x)\equiv 0$, $r(x)$ and $q(x)$ are completely monotone on the half-line $x>0$ (that is, $z>1$) if $n^2-\frac{1}{4}\geqq 0$. In fact, since

$$(12)\qquad \cosh x/\sinh x=(1+e^{-2x})/(1-e^{-2x})=(1+e^{-2x})\sum_{j=0}^{\infty}e^{-2jx}$$

holds for $x>0$, it is seen that $r(x)$ is the sum of completely monotone functions and is therefore completely monotone. Similarly, the identity

$$(13)\qquad 1/\sinh x=2e^{-x}/(1-e^{-2x})=2e^{-x}\sum_{j=0}^{\infty}e^{-2jx}$$

shows that $1/\sinh x$ is completely monotone. Since the product of completely monotone functions is completely monotone, it follows that $1/\sinh^2 x$ is. Hence, $q(x)$ is completely monotone for $x>0$ if $n^2-\frac{1}{4}\geqq 0$. The first assertion of (I) follows from the theorem concerning (2), quoted in the Introduction.

By (11), the inequality $n^2-\frac{1}{4}>0$ implies that $q(x)$ satisfies (7); so that (10) has only one solution, up to a constant factor, representable in the form (4). If $n^2-\frac{1}{4}=0$ and $m>0$, the general solution of (10) is

$$y(x)=c_1\exp(m\int_1^x ds/\sinh s)+c_2\exp(-m\int_1^x ds/\sinh s).$$

It is readily verified that this solution $y(x)$ is completely monotone for

$0 < x < \infty$ if and only if $| c_1 | \leqq c_2 \exp(-2m \int_1^\infty ds/\sinh s)$. Finally, if $n^2 - \frac{1}{4} = 0$ and $m = 0$, the general solution of (10) is $y(x) = c_1 \int_1^x ds/\sinh s + c_2$, which is completely monotone if and only if $0 \leqq -c_1 \leqq c_2 / \int_1^\infty ds/\sinh s$. This completes the proof of (I).

Remark. Since the change of variables $z \rightarrow -z$ leaves (9) unaltered, it follows that if $\nu(\nu+1) \geqq 0$, there exists on $-\infty < z < -1$ a solution $y = y(z)$ of (9) representable in the form (4), where $z = -\cosh x$.

2. Hypergeometric functions. The case $\mu = 0$ of (8) can be considered as a special case of the hypergeometric differential equation

$$(14) \qquad z(1-z)d^2y/dz^2 + [c - (a+b+1)z]dy/dz - aby = 0.$$

In fact, if the change of independent variables $2z - 1 \rightarrow z$ is made in (14) and if $c = 1$, $a = \nu + 1$ and $b = -\nu$, then (14) becomes (8) with $\mu = 0$. Thus the results of Section 1 can be transferred from (9) to (14) if

$$(15) \qquad c = 1, \qquad a + b = 1, \qquad ab \leqq 0.$$

But it turns out the corresponding statements are valid for a wider range of parameters than those specified by (15).

(II) *If the parameters satisfy*

$$(16) \qquad ab \leqq 0 \text{ and } a + b \geqq \max(2c - 1, 0),$$

then (14) *possesses a solution* $y = y(z)$ *representable in the form* (4), *where* $0 < x < \infty$ *and either* $2z - 1 = \cosh x$ (*so that* $1 < z < \infty$) or $2z - 1 = -\cosh x$ (*so that* $-\infty < z < 0$). *This solution of* (14) *is unique, up to a constant factor, if and only if either* $ab < 0$ *or* $a = b = 0$.

If x is eliminated, it is seen that (II) means that (14) has, under the assumption (16), a solution of the form

$$(17) \qquad y(z) = \int_0^\infty [2z - 1 + 2(z^2 - z)^{\frac{1}{2}}]^{-t} d\phi(t)$$

for $1 < z < \infty$, or of the form

$$\text{(17 bis)} \qquad y(z) = \int_0^\infty [1 - 2z + 2(z^2 - z)^{\frac{1}{2}}]^{-t} d\phi(t)$$

for $-\infty < z < 0$, where $\phi(t)$ is monotone on $[0, \infty)$ and not a constant.

Proof of (II). Introduce x as a new independent variable in (14), where either $2z - 1 = \cosh x$ or $2z - 1 = -\cosh x$. The resulting differential equation in either case is

$$\text{(18)} \quad d^2y/dx^2 + [(a + b)\cosh x/\sinh x + (a + b + 1 - 2c)/\sinh x] dy/dx + aby = 0.$$

A comparison of this differential equation with (2) shows that $p(x) \equiv 1$, $r(x) = (a + b)\cosh x/\sinh x + (a + b + 1 - 2c)/\sinh x$ and $q(x) = -ab$. It follows therefore from (12) and (13) that

$$r(x) = (a + b)(1 + 2 \sum_{j=1}^{\infty} e^{-2jx}) + 2(a + b + 1 - 2c) \sum_{j=0}^{\infty} e^{-(2j+1)x}.$$

Thus $r(x)$ is completely monotone on $(0, \infty)$ if and only if $a + b \geqq 0$ and $a + b + 1 - 2c \geqq 0$; that is, if and only if $a + b \geqq \max(0, 2c - 1)$. Clearly, $q(x)$ is completely monotone when $ab \leqq 0$. Hence, the first assertion of (II) follows from the theorem concerning (2), quoted in the Introduction.

If $ab < 0$, then $q(x)$ satisfies (7), hence (17) has at most one solution $y(x)$ representable in the form (4). If $ab = 0$, the general solution of (17) is

$$y(x) = c_1 \int_1^x (\sinh s)^{-(a+b)} \exp[(2c - a - b - 1) \int_1^s dt/\sinh t] ds + c_2.$$

If $a + b > 0$, it is easily seen that $y(x) \geqq 0$ for $0 < x < \infty$ if and only if

$$0 \leqq -c_1 \leqq c_2 / \int_1^\infty (\sinh s)^{-(a+b)} \exp[(2c - a - b - 1) \int_1^s dt/\sinh t] ds;$$

in which case $y(x)$ is completely monotone. But if $a + b = 0$, then $y(x) > 0$ for $0 < x < \infty$ if and only if $c_1 = 0 \leqq c_2$. Thus the second part of (II) follows.

Integral representations distinct from (17), (17 bis) can be obtained for certain solutions of (14). In this direction, the following will be proved:

(III) *If the parameters satisfy*

$$1 + c(a + b) \geqq c + \tfrac{1}{2}(a + b)^2, \tag{19}$$

then (14) *has a solution representable in the form*

$$y(t) = (-z)^{\frac{1}{2}(1-c)}(1-z)^{\frac{1}{2}(c-a-b-1)} \int_0^\infty d\phi(t)/(1-z)^t \tag{20}$$

for $-\infty < z < 0$, *while if*

$$c + \tfrac{1}{2}(a + b)^2 \geqq c(a + b), \tag{21}$$

then (14) *has a solution representable in the form*

$$y(z) = z^{-\frac{1}{2}c}(z-1)^{\frac{1}{2}(c-a-b)} \int_0^\infty d\phi(t)/z^t \tag{22}$$

for $1 < z < \infty$, *where* $0 \not\equiv d\phi(t) \geqq 0$ *in both cases.*

Proof of (III). For $-\infty < z < 0$, the variation of constants and the change of independent variables determined by

$$u = (-z)^{\frac{1}{2}(c-1)}(1-z)^{\frac{1}{2}(a+b+1-c)}y, \qquad x = \log(1-z) \tag{23}$$

transform (14) into a differential equation, for u, of the form (2), where

$$p(x) = 1 - e^{-x}, \qquad r(x) = e^{-x} \tag{24}$$

and

$$4q(x) = (1-c)^2/(e^x - 1) + [1 - (a + b - c)^2]/e^x + (a - b)^2;$$

cf. [4], p. 453. Clearly, $p(x) > 0$, and $p'(x) = r(x) = e^{-x}$ is completely monotone for $0 < x < \infty$. Since $1/(e^x - 1) = e^{-x}/(1 - e^{-x}) = \sum_{j=1}^{\infty} e^{-jx}$ for $0 < x < \infty$, it follows that

$$4q(x) = (a - b)^2 + [(1-c)^2 + 1 - (a + b - c)^2]e^{-x} + (1-c)^2 \sum_{j=2}^{\infty} e^{-jx}.$$

Thus $q(x)$ is completely monotone for $0 < x < \infty$ if and only if the coefficient of e^{-x} is non-negative; that is, if and only if (19) holds. If (19) is satisfied, then the differential equation for u, which resulted from (14), has a solution of the form (4), that is,

$$u(x) = \int_0^\infty e^{-xt} d\phi(t), \text{ where } 0 \not\equiv d\phi(t) \geqq 0 \qquad (0 < x < \infty). \tag{25}$$

In view of (23), this is equivalent to (20).

Similarly, if $1 < z < \infty$, the variation of constants and the change of independent variables determined by

$$u = z^{\frac{1}{2}c}(z-1)^{\frac{1}{2}(a+b-c)}y, \qquad x = \log z$$

transform (14) into a differential equation, for u, of the form (2), where (24) holds and

$$4q(x) = (a+b-c)^2/(e^x-1) + (2c-c^2)/e^x + (a-b)^2.$$

The remainder of the proof of the second part of (III) is similar to the proof of the first part.

The last two theorems concern certain solutions of the hypergeometric differential equation (14) on the real z-range between the singular points $-\infty, 0$ or between the singular points $1, \infty$. In some cases, it is possible to ascertain the complete monotony of solutions of (14) on the part of the interval between the singular points $0, 1$.

(IV) *If there exists a number* α *satisfying*

$$(26) \qquad 0 < \alpha \leqq \tfrac{1}{2}, \qquad \alpha(a+b+1) \leqq c,$$

and if

$$(27) \qquad ab \geqq 0, \qquad a+b+1 \geqq 0,$$

then (14) *has a solution* $y = y(z) \not\equiv 0$ *which is completely monotone on the interval* $0 < z < \alpha$.

Proof of (IV). The coefficient, $P(z) = z(1-z) = z - z^2$, of d^2y/dx^2 in (14) is non-negative and possesses a completely monotone derivative, $P'(z) = 1 - 2z$, on the interval $(0, \frac{1}{2})$. The coefficient, $Q(z) = ab$, of $-y$ in (14) is completely monotone if $ab \geqq 0$. Finally, the coefficient, $R(z) = c - (a+b+1)x$, of dy/dx is completely monotone on $(0, \alpha)$ if the second inequality holds in both (26) and (27). In view of the theorem concerning (2), quoted in the Introduction, (IV) follows.

It should be mentioned that the result is not a consequence of standard integral representations of certain hypergeometric functions (as given, for example, by the formula near the bottom of p. 8 in [2]).

3. Confluent hypergeometric functions. Consider Kummer's differential equation,

$$(28) \qquad z\,d^2y/dz^2 + (c-z)\,dy/dz - ay = 0.$$

A standard integral representation (cf. the last formula in [2], p. 87) shows that if

$$c > a > 0, \tag{29}$$

then (28) possesses a solution of the form (4), where $z = -x$. It turns out that (28) possesses such solutions for parameter ranges different from those specified by (29).

(V) *If the parameters satisfy*

$$a \leqq 0 \text{ and } c \geqq 0, \tag{30}$$

then (28) *possesses a solution representable in the form* (4) *for* $1 < x < \infty$, *where* $z = -x$ (*so that* $-\infty < z < 0$), *and this solution is unique, up to a constant factor, if and only if* $a < 0$; *furthermore, if*

$$0 \leqq c \leqq 2a, \tag{31}$$

then (28) *possesses a solution representable in the form*

$$y(z) = e^{\frac{1}{2}z} \int_0^\infty e^{-zt} d\phi(t), \text{ where } 0 \not\equiv d\phi(t) \geqq 0 \qquad (0 < z < \infty), \tag{32}$$

and this solution is unique up to a constant factor.

For the case in which Kummer's normal form, (28), is replaced by that of Whittaker, cf. [6].

Proof of (V). In order to prove the first part of (V), apply the change of independent variables $z \to -x$ in (28) and write the resulting differential equation in the form

$$d^2y/dx^2 + (1 + c/x)dy/dx + (a/x)y = 0. \tag{33}$$

A comparison of (33) with (2) shows that $p(x) \equiv 1$, $r(x) = 1 + c/x$, $q(x) = -a/x$; so that $p > 0$, and p', r, q are completely monotone for $0 < x < \infty$, if (30) holds. Hence (20) has a non-trivial solution representable, for $0 < x < \infty$, in the form (4), where $z = -x$.

In order to ascertain the uniqueness of the solution when $a < 0$, it is sufficient to verify that (6) is violated in this case. But (6) becomes

$$\text{const.} \int_1^\infty e^{-s} s^{-c} \left(\int_1^s e^t t^{c-1} dt \right) ds < \infty, \qquad (X = 1),$$

and this cannot be satisfied, since the interior integral is asymptotically proportional to $e^s s^{c-1}$, as $s \to \infty$. If $a = 0$, the general solution of (33) is

$$y(x) = c_1 \int_1^x e^{-s} s^{-c} ds + c_2,$$

from which it is easily seen that $y(x)$ is completely monotone if $c_1 \leqq 0$ and if c_2 is so chosen that $y(\infty) \geqq 0$.

In order to prove the statements concerning (31) and (32), use will be made of the following variation of constants:

$$y = e^{\frac{1}{2}z} u \qquad (z = x). \tag{34}$$

This transforms (28) into

$$d^2u/dx^2 + (c/x)\,du/dx - [\tfrac{1}{4} + (2a - c)/2x]u = 0.$$

Clearly, if (31) holds, this differential equation possesses a solution representable, for $0 < x < \infty$, in the form (25). This solution is unique, up to a constant factor, since $q(x)$ satisfies (7). Finally, it follows from (34) and (25) that (20) possesses a solution of the form (32).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman and A. Wintner, "On the Laplace-Fourier transcendents," *American Journal of Mathematics*, vol. 71 (1949), pp. 367-372.

[2] W. Magnus and F. Oberhettinger, *Formulas and theorems for the special functions of mathematical physics*, New York, 1949.

[3] H. Weyl, "Ueber gewöhnliche lineare Differentialgleichungen mit singulären Stellen und ihre Eigenfunktionen," *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, 1909, pp. 37-63.

[4] ———, "Ueber gewöhnliche lineare Differentialgleichungen mit singulären Stellen und ihre Eigenfunktionen, II," ibid., 1910, pp. 442-467.

[5] A. Wintner, "On the Laplace-Fourier transcendents occurring in mathematical physics," *American Journal of Mathematics*, vol. 69 (1947), pp. 87-98.

[6] ———, "On the Whittaker functions $W_{km}(x)$," *Journal of the London Mathematical Society*, vol. 25 (1950), pp. 351-353.

ON THE NON-INCREASING SOLUTIONS OF $y'' = f(x, y, y')$.*

By PHILIP HARTMAN and AUREL WINTNER.

1. A theorem of A. Kneser [7], pp. 183-191, can be formulated as follows: Suppose that $f(x, y)$ is defined for $0 \leqq x < \infty$, $0 \leqq y < \infty$ in such a way that $f(x, y)$ is continuous; that $f(x, y) \geqq 0$; that $f(x, 0) \equiv 0$ for $0 \leqq x < \infty$; that $f(x, y)$ satisfies, with respect to y, a Lipschitz condition which is uniform in every (x, y)-rectangle; finally, that $f(x, y)$ is monotone with respect to y for every fixed $x \geqq 0$. Then, for every $y_0 > 0$, the differential equation

$$y'' = f(x, y) \tag{1}$$

has a solution $y = y(x)$ which belongs to the initial condition

$$y(0) = y_0, \tag{2}$$

exists on the half-line $0 \leqq x < \infty$, and satisfies the inequalities

$$y(x) \geqq 0 \text{ and } y'(x) \leqq 0. \tag{3}$$

It follows from a remark of Mambriani [8], p. 622, that the condition that $f(x, y)$ be monotone with respect to y can be omitted in this theorem.

The most immediate application of Kneser's theorem is the linear case,

$$y'' = q(x)y, \tag{4}$$

of (1). The above conditions are satisfied by $f(x, y) = q(x)y$, if $q(x)$ is a continuous, non-negative function for $0 \leqq x < \infty$. From this case, (4), of Kneser's theorem, it is easy to deduce a somewhat more general result; namely, that if $q(x)$ is a continuous, non-negative function for $0 \leqq x < \infty$ and $r(x)$ is an arbitrary continuous function for $0 \leqq x < \infty$, then the linear differential equation

$$y'' = r(x)y' + q(x)y \tag{5}$$

has, for every $y_0 > 0$, a solution $y = y(x)$ satisfying (2) and (3); cf. [5], p. 368.

There naturally arises the question as to whether or not Kneser's theorem

---

* Received July 2, 1950

can be generalized so as to apply to a non-linear differential equation of the form

$$(6)\qquad y'' = f(x, y, y').$$

Such generalizations are known; cf., e. g., [11], p. 623; [12], p. 105. None of these results is, however, sufficiently general to include directly the special case (5) of (6). A general theorem of this type, to be proved below, is as follows:

THEOREM I. *Let* $f(x, y, z)$ *be a continuous function on the octant*

$$(7)\qquad R\colon\ 0 \leqq x < \infty, \qquad 0 \leqq y < \infty, \qquad -\infty < z \leqq 0$$

and suppose that

$$(8)\qquad f(x, 0, 0) \equiv 0 \qquad \text{for } 0 \leqq x < \infty;$$

that

$$(9)\qquad f(x, y, 0) \geqq 0 \ \text{ for } 0 \leqq x < \infty,\ 0 \leqq y < \infty;$$

finally that, for every $C > 0$, *there exists a positive continuous function* $\phi(z) = \phi_C(z)$ *for* $-\infty < z \leqq 0$, *satisfying*

$$(10)\qquad |f(x, y, z)| \leqq \phi(z) \ \text{ for } 0 \leqq x \leqq C, \qquad 0 \leqq y \leqq C, \qquad -\infty < z \leqq 0$$

and

$$(11)\qquad \int_{-\infty}^{0} z\,dz/\phi(z) = -\infty.$$

Then there belongs to every $y_0 > 0$ *a solution* $y = y(x)$ *of* (6) *which exists for* $0 \leqq x < \infty$ *and satisfies* (2) *and* (3).

2. The comparatively simple nature of the proof of Theorem I will be due to a consistent application of certain results of Nagumo [9]. The proofs of Theorem I and of Theorem III below are very similar to the geometrical qualitative arguments employed in [4].

In order to prove Theorem I, a few general lemmas will be needed.

LEMMA 1. *Let* $f(x, y, z)$ *be a continuous function on* $0 \leqq x < \infty$, $0 \leqq y < \infty$, $-\infty < z < \infty$, *satisfying*

$$(12)\qquad f(x, y, 0) > 0 \ \text{ for } 0 \leqq x < \infty, \qquad 0 < y < \infty.$$

Let $y = y(x)$ *be a solution of* (1) *on an interval* $(0 \leqq)\ a \leqq x \leqq b$, *belonging to initial conditions satisfying*

$$(13)\qquad y(a) > 0 \ \text{ and } \ y'(a) \geqq 0.$$

Then

$$y'(x) > 0 \text{ for } a < x \leqq b. \tag{14}$$

Proof of Lemma 1. The argument used in the proof will be similar to one in [13], p. 46.

It will first be shown that there cannot exist a value of x, say $x = \xi$, such that $a < \xi < b$, $y'(\xi) < 0$ (and $y(\xi) > 0$). For if such an x-value exists, let $x = \alpha (\geqq a)$ be the greatest value of x to the left of $x = \xi$ for which $y'(x) = 0$. Clearly, $y'(x) < 0$ for $\alpha < x \leqq \xi$. Hence, $y(\alpha) > y(\xi) > 0$. Consequently, (6) and (12) imply the inequality $y''(\alpha) = f(\alpha, y(\alpha), 0) > 0$. But this, together with $y'(\alpha) = 0$, contradicts $y'(x) < 0$ for $\alpha < x \leqq \xi$.

This proves a weakened form of (14),

$$y'(x) \geqq 0 \text{ for } a \leqq x \leqq b. \tag{15}$$

Since (12) implies that $y \equiv y(a)$ cannot be a solution of (6), it follows that $y'(a_1) > 0$ for some $x = a_1 (> a)$ arbitrarily near $x = a$. Arguing as above, it is seen that there cannot exist a value of $x = \xi > a_1$, where $y'(\xi) = 0$. This proves Lemma 1.

3. It is interesting to note that if (12) is weakened to (9), then Lemma 1 becomes false even if (14) is weakened to (15). This is shown by the example, $f(x, y, z) \equiv 2y \mid z \mid^{\lambda} \operatorname{sgn} z/(2 - \lambda)$, where $0 < \lambda < 1$. Correspondingly, (6) becomes

$$y'' = 2y \mid y' \mid^{\lambda} \operatorname{sgn} y'/(2 - \lambda). \tag{16}$$

If (16) is multiplied by $\mid y' \mid^{1-\lambda}(2 - \lambda)$, a quadrature gives

$$\mid y' \mid^{2-\lambda} = (y^2 - c) \operatorname{sgn} y',$$

where $c > 0$ is an integration constant. Accordingly,

$$y' = \pm \{(y^2 - c) \operatorname{sgn} y'\}^{1/(2-\lambda)},$$

where both signs, ±, are admissible. If the minus sign is chosen, a quadrature gives a solution of (16), satisfying $y(0) = c^{\frac{1}{2}} > 0$, $y'(0) = 0$, but $y'(x) < 0$ for small $x > 0$.

A criterion which prevents such a possibility is as follows:

LEMMA 2. *Let $f(x, y, z)$ be a continuous function on $0 \leqq x < \infty$, $0 \leqq y < \infty$, $-\infty < z < \infty$, satisfying* (8) *and* (9). *In addition, for every $C > 0$, let there exist an $M = M_C$ such that*

$$f(x, y, z) - f(x, y, 0) \geqq Mz \text{ for } 0 \leqq x \leqq C,\ 0 \leqq y \leqq C,\ -C \leqq z \leqq C.$$

Let $y(x)$ be the same as in Lemma 1. *Then* (15) *holds. If $y'(a) > 0$, then* (14) *holds.*

Proof of Lemma 2. Suppose that (15) does not hold. Then there exists a value of x, say $x = \xi$, such that $a < \xi \leqq b$, $y(\xi) > 0$ and $y'(\xi) < 0$. Let $x = \alpha (\geqq a)$ be the greatest value of x to the left of $x = \xi$ for which $y'(x) = 0$. For $\alpha < x \leqq \xi$, put

$$q(x) = \{f(x, y(x), 0) - f(x, 0, 0)\}/y(x) \tag{17}$$

and

$$r(x) = \{f(x, y(x), y'(x)) - f(x, y(x), 0)\}/y'(x). \tag{18}$$

Then (6) shows that $y(x)$ is a solution of (5) for $\alpha < x \leqq \xi$. It follows from (8) and (9) that $q(x) \geqq 0$. The condition on f involving the constant M implies that $r(x)$ is bounded from above by M for $\alpha < x \leqq \xi$.

If a new independent variable $s = s(x)$, defined by

$$ds = \exp(-\int_x^\xi r(u)\,du)\,dx,$$

is introduced, then (5) becomes

$$d^2y/ds^2 = q(x)y \exp 2\int_x^\xi r(u)\,du,$$

where $x = x(s)$. Since the coefficient of y is non-negative, it follows that $y(x)$, as a function of s, is convex upwards. Hence dy/ds is a non-decreasing function of s. In terms of x, this means that

$$y'(x) \exp \int_x^\xi r(u)\,du \leqq y'(\xi) < 0$$

for $\alpha < x \leqq \xi$. Hence, by the boundedness of $r(x)$ from above and by the continuity of $y'(x)$,

$$y'(\alpha) \exp M(\xi - \alpha) < 0.$$

But this contradicts the fact that $y'(\alpha) = 0$. This proves the first assertion of Lemma 2, and the second assertion is proved in a similar manner.

4. The example (16) shows that Lemma 2 becomes false if the one-sided Lipschitz condition at $z = 0$ is weakened to a Hölder condition and the other assumptions are retained.

Lemma 1 corresponds to a uniqueness theorem ([13], p. 45) which states that if, on the (x, y, z)-space, $f(x, y, z)$ is a continuous function which is strictly increasing with y (for fixed x, z), then (6) has at most one solution joining two distinct points of the (x, y)-plane. The condition that $f(x, y, z)$ is "increasing" with y cannot be lightened to "non-decreasing" with y; cf. [13], p. 45. A simple calculation shows that if the factor sgn y' is omitted, then (16) also becomes a counter-example. What corresponds to Lemma 2 is the following uniqueness theorem:

THEOREM II. *If $f(x, y, z)$ is a continuous function on the (x, y, z)-space, is non-decreasing with respect to y (for fixed x, z) and satisfies, with respect to z, a Lipschitz condition which is uniform on every bounded portion of the (x, y, z)-space, then* (6) *has at most one solution $y = y(x)$ joining two distinct points of the (x, y)-plane.*

The proof is similar to that of Lemma 2 and will be omitted. Under the additional hypothesis that $f(x, y, z)$ has continuous partial derivatives with respect to y and z, the assertion of Theorem II was proved in [10], p. 105 and [2], p. 20.

5. For the proof of Theorem I, one more lemma will be needed.

LEMMA 3. *Let $f(x, y, z)$ satisfy the conditions of Theorem* I. *Let $x_0 > 0$ and $y_0 > 0$. Then* (6) *has on $0 \leqq x \leqq x_0$ a solution $y = y(x)$ satisfying*

$$y(0) = y_0 \text{ and } y(x_0) = 0. \tag{19}$$

Proof of Lemma 3. Let the definition of $f(x, y, z)$ be extended by placing $f(x, y, z) = f(x, y, 0)$ for $z > 0$; so that $f(x, y, z)$ is a continuous function for $0 \leqq x < \infty$, $0 \leqq y < \infty$, $-\infty < z < \infty$. The resulting differential equation (6) has a solution satisfying (19).

The last assertion follows from a result of Nagumo [9], p. 864, if slight additional conditions are imposed in $f(x, y, z)$; namely, that condition (9) be replaced by (12) and that $f(x, y, z)$ have continuous partial derivatives with respect to y and z. (The wording of Nagumo's theorem, but not the proof, requires that (8) be replaced by $f(x, 0, 0) < 0$). Actually, these additional conditions are not needed; see [14], pp. 256-258; [3], pp. 277-278 and [19], pp. 55-56. (Incidentally, the superfluity of these additional conditions can be shown by using Nagumo's result and by approximating $f(x, y, z)$ suitably, and Nagumo's Hilfssatz 1 and 2 show that these approximations need only be carried out on some parallelepiped of the type $0 \leqq x \leqq x_0$, $0 \leqq y \leqq y_0$, $-C \leqq z \leqq C$; cf. the considerations below.)

Let $y = y(x)$ be a solution of (6) satisfying (19). In order to complete the proof of Lemma 3, it remains to show that $y(x)$ can be chosen so as to satisfy $y'(x) \leqq 0$ for $0 \leqq x \leqq x_0$. This will prove that the extension of the definition of $f(x, y, z)$ was unnecessary.

The inequality $y'(x) \leqq 0$ is clear if $f(x, y, z)$ satisfies (12). For, by Lemma 1, $y'(x) < 0$ whenever $y(x) > 0$.

If $f(x, y, z)$ does not satisfy (12), consider the differential equation

$$y'' = f(x, y, y') + \epsilon y \qquad (\epsilon > 0) \tag{6 bis}$$

instead of (6). The right-hand side of (6 bis) corresponds to the function $g(x, y, z) = f(x, y, z) + \epsilon y$. This function satisfies the analogue of (8), namely, $g(x, 0, 0) \equiv 0$; and the analogue of (12), namely, $g(x, y, 0) > 0$ for $0 \leqq x < \infty$, $0 < y < \infty$. In (10), let $C = \max(x_0, y_0)$, and let $\phi(z) = \phi_C(z)$ be defined as $\phi(0)$ for $z > 0$. Clearly,

$$\int_{-\infty} z dz/\phi(z) = -\infty \text{ and } \int^{\infty} z dz/\phi(z) = \infty,$$

while $|g(x, y, z)| \leqq \phi(z) + \epsilon y_0$ for $0 \leqq x \leqq x_0$, $0 \leqq y \leqq y_0$, $-\infty < z < \infty$. It is also clear that there exists a pair of positive constants ϵ_0, M for which

$$\int_{-M}^{-a} z dz/(\phi(z) + \epsilon y_0) > 1 + y_0 \text{ and } \int_{a}^{M} z dz/(\phi(z) + \epsilon y_0) > 1 + y_0$$

whenever $0 < \epsilon \leqq \epsilon_0$, where $a = 1 + 2y_0/x_0$. It follows from the proof of Nagumo's Satz 2, [9], pp. 864-865, that if $0 < \epsilon \leqq \epsilon_0$, then (6 bis) possesses a solution $y = y_\epsilon(x)$ satisfying (19), provided that $g(x, y, z)$ possesses continuous partial derivatives with respect to y and z. Now, $y_\epsilon'(x) \leqq 0$, by Lemma 1.

Actually, this solution satisfies

$$0 \geqq y_\epsilon'(x) \geqq -M \text{ for } 0 \leqq x \leqq x_0 \text{ and } 0 < \epsilon \leqq \epsilon_0. \tag{20}$$

For suppose, if possible, that $y_\epsilon'(x) < -M$ for some value of x, say $x = \xi$. Consider the two cases $0 \leqq \xi \leqq \frac{1}{2}x_0$ and $\frac{1}{2}x_0 < \xi \leqq x_0$. In the first case, there exists an x-value $x = x_1$ such that $y_\epsilon(x_1) = 0$, $y_\epsilon'(x_1) < 0$ and $\xi \leqq x_1 < \xi + \frac{1}{2}x_0 (\leqq x_0)$; cf. Hilfssatz 2, [9], p. 863. But then $y_\epsilon(x)$ does not exist for $x_1 \leqq x \leqq x_0$, since $f(x, y, z)$ is only defined for $y \geqq 0$. In the second case, $\frac{1}{2}x_0 < \xi \leqq x_0$, there exists an x-value $x = x_1$ such that $y_\epsilon(x_1) = y_0$, $y_\epsilon'(x_1) < 0$ and $(0 \leqq)\ \xi - \frac{1}{2}x_0 < x_1 \leqq \xi$; cf. Hilfssatz 2, [9], p. 863. But $y_\epsilon'(x) \leqq 0$ for $0 \leqq x \leqq x_0$, so this contradicts $y_\epsilon(0) = y_0$.

Thus the proviso that $g(x, y, z)$ have continuous partial derivatives with respect to y and z can be dropped. This follows from the general theorems quoted above; for, although the majorant function $\phi(z) + \epsilon y_0$ need not satisfy

$$\int_{-\infty} z\,dz/(\phi(z) + \epsilon y_0) = -\infty,$$

the considerations above show that it is sufficient to consider $g(x, y, z)$ on the parallelepiped $0 \leqq x \leqq x_0$, $0 \leqq y \leqq y_0$, $-M \leqq z \leqq M$. Alternatively, as mentioned above, it is sufficient to use the Nagumo theorem and to approximate $g(x, y, z)$ suitably on this parallelepiped.

The differential equation (6 bis) and the inequalities (20) show that, by equicontinuity, there exists a sequence $\epsilon_0 > \epsilon_1 > \epsilon_2 \cdots$ having the property that $y_\epsilon(x)$, where $\epsilon = \epsilon_n \to 0$, tends to a solution $y(x)$ of (6), as $n \to \infty$. Clearly, $y = y(x)$ satisfies (19) and $y'(x) \leqq 0$ for $0 \leqq x \leqq x_0$. This proves Lemma 3.

6. *Proof of Theorem* I. Let $y_0 > 0$ be fixed and let $y = y(x, x_0)$ be a solution of (6), on $0 \leqq x \leqq x_0$, satisfying (19). If there exists, for some x_0, an x-value, say $x = \xi$, such that $0 < \xi < x_0$ and $y(\xi, x\,) = 0$, then it follows from $y'(x, x_0) \leqq 0$ that $y'(x, x_0) \equiv 0$ for $\xi \leqq x \leqq x_0$. Thus the solution $y(x, x_0)$ can be extended for $x_0 \leqq x < \xi$ by placing $y(x, x_0) \equiv 0$ for $x > x_0$. Hence $y = y(x, x_0)$ is a solution of (6) for $0 \leqq x < \infty$ and satisfies (2) and (3).

Consequently, it can be supposed that $y(x, x_0) > 0$ for $0 \leqq x < x_0$. Choose $x_0 = n$, where $n = 1, 2, \cdots$. It then follows from Nagumo's Hilfssatz 2 ([9], p. 863) that, for any integer $k > 0$, there exists a constant $M = M(k) > 0$ such that

$$0 \geqq y'(x, n) \geqq -M \text{ for } 0 \leqq x \leqq k \text{ and } n = k + 1, k + 2, \cdots.$$

By equicontinuity, there exists an increasing sequence of integers $n_1, n_2, \cdots$ such that $y(x, n_1), y(x, n_2), \cdots$ tends to a solution $y = y(x)$ of (6) for $0 \leqq x < \infty$. Clearly, $y = y(x)$ satisfies (2) and (3). This proves Theorem I.

7. The indispensability of Nagumo's condition, (10)-(11), in Theorem I is shown by the following example: Let $f(x, y, x) = 3(1 + z^2)^{1+\lambda}y^2/2\lambda$, where $0 < \lambda < 1$. Then (6) becomes

$$y'' = 3(1 + y'^2)^{1+\lambda}y^2/2\lambda. \tag{21}$$

If (21) is multiplied by $2\lambda y'/(1+y'^2)^{1+\lambda}$ a quadrature gives

$$1/(1+y'^2)^{\lambda} = c - y^3,$$

where c is an integration constant. Consider a solution $y = y(x)$ of (21) satisfying $y(0) = y_0 > 0$. For such a solution of (21), the constant c satisfies $y_0{}^3 < c \leqq 1 + y_0{}^3$. Let $y_0 > 1$, so that $c - 1 > 0$. The equation in the last formula line can be written in the form

$$y' = -\{1 - (c - y^3)^{1/\lambda}\}^{\frac{1}{2}}/(c - y^3)^{\frac{1}{2}\lambda},$$

if $y' \leqq 0$. Thus, if $y'(0) < 0$, the solution $y(x)$ continues to decrease as x increases, unless $y^3(x)$ becomes $c - 1$ for some finite value of x, say $x = \xi$.

If such a value of $x = \xi$ exists, then, by (21) and a convexity argument, $y'(x) > 0$ must hold for those $x > \xi$ for which $y(x)$ exists. But if $y_0 > 1$, and if c is an admissible integration constant, it follows from (21) and

$$\int_a^b (c - y^3)^{\frac{1}{2}\lambda}/\{1 - (c - y^3)\}^{\frac{1}{2}}\, dy < \infty, \text{ where } a^3 = c - 1 \text{ and } b^3 = c,$$

that there exists a $\xi = \xi(y_0, c)$ such that $y^3(\xi) = c - 1$. Consequently, if $y_0 > 1$, then (21) does not have a solution on $0 \leqq x < \infty$ satisfying (2) and (3).

8. There arises the need for a criterion which takes into account the complication presented by the counter-example (21). Such a criterion can be formulated as follows:

THEOREM III. *Let $f(x, y, z)$ be defined and continuous on a domain containing the octant* (7) *in its interior. Suppose that the solutions of* (6) *are uniquely determined* (*locally*) *by their initial conditions, that* (8) *holds, and that*

$$f(x, y, z) \geqq 0 \text{ on } R. \tag{22}$$

Then there exists a positive $Y\,(\leqq \infty)$ having the property that if $0 < y_0 < Y$, then (6) *possesses a solution on $0 \leqq x < \infty$ satisfying* (2) *and* (3)

Proof of Theorem III. Let $x_0 > 0$ be fixed and consider the solutions $y = y(x; \mu)$ of (6) determined by the initial conditions $y(x_0) = 0$ and $y'(x_0) = \mu \leqq 0$. Since (8) implies that $y(x, 0) \equiv 0$, it follows from the continuous dependence of solutions on initial conditions that, if $|\mu|$ is sufficiently small, $y(x; \mu)$ exists for $0 \leqq x \leqq x_0$. Furthermore, (22) implies that $y(x; \mu)$ is convex upwards if $\mu < 0$. Hence, $y'(x; \mu) \leqq \mu < 0$ for

$0 \leqq x \leqq x_0$. Consequently, $y(0;\mu) > 0$ if $\mu < 0$. In addition, $y(0;\mu)$ is a continuous function of μ. Hence, there exists a Y, where $0 < Y \leqq \infty$, with the property that $y_0 = y(0;\mu)$ for some μ if $0 < y_0 < Y$. In what follows, it will be supposed that y_0 has been chosen in this fashion.

Consider the solution $y = y(x, m)$ of (6) determined by the initial condition $y(0) = y_0$, $y'(0) = m$. Let Ω denote the set of negative m-values for which the solution $y = y(x, m)$ can be extended over some interval $0 \leqq x \leqq x_1$ and satisfies $y(x_1) = 0$. By the choice of y_0, the set Ω is not empty. In fact, Ω is an open set on the m-axis. For, if $y = y(x) = y(x, m)$ satisfies $y(x_1) = 0$, then $y'(x_1, m) < 0$, since $y \equiv 0$ is the only solution of (6) satisfying $y(x_0) = 0$, $y'(x_0) = 0$. It follows that, if m belongs to Ω, all numbers sufficiently close to m belong to Ω.

Let $y_0'(\leqq 0)$ be the least upper bound of Ω. It will be shown that $y = y(x) = y(x, y_0')$ exists for $0 \leqq x < \infty$ and satisfies (2) and (3).

First, if $y_0' = 0$, then $y(x)$ cannot vanish, since (22) implies $y''(x) \geqq 0$. If $y_0' < 0$, then $y(x)$ cannot vanish. For otherwise there exist values of $m > y_0'$ for which $y(x, m)$ has a zero, and this contradicts the definition of y_0'. (In particular, y_0' is a cluster point of Ω.)

Next, $y'(x)$ cannot attain a positive value for some value of x, say for $x = \xi$. For if this were possible, it would follow that there exist in Ω values of m for which $y(x, m) > 0$ for $0 \leqq x \leqq \xi$, and $y'(\xi, m) > 0$, while $y(x, m)$ vanishes for some $x_1 > \xi$. But this contradicts the inequality $y'(x, m) \leqq 0$, where $0 \leqq x \leqq x_1$. Hence, $y(x)$ satisfies (2) and (3) on its domain of existence.

The solution $y = y(x)$ can fail to exist for all $x \geqq 0$ only if there exists a positive number $x = x_2$ such that $|y(x)| + |y'(x)| \to \infty$ as $x \to x_2 - 0$; cf. [16]. But this is impossible since $0 \leqq y(x) \leqq y_0$; $y_0' \leqq y'(x) \leqq 0$, by virtue of $y''(x) \geqq 0$. This proves Theorem III.

9. The proofs of Theorem I and III seem to indicate that, if $f(x, y, z)$ is continuous and non-negative on R, and if (6) possesses a solution on $0 \leqq x < \infty$ satisfying (2) and (3), then the points $(0, y_0)$ and $(x_0, 0)$, where $x_0 > 0$, can be joined by a solution of (6). But it turns out that such is not the case.

In fact, consider the differential equation

$$y'' = -(1 - y')^2 y'. \tag{23}$$

The function $f(x, y, z)$ corresponding to (23) is $f(x, y, z) = -(1 - z)^2 z$, which is non-negative on R. For every y_0, the differential equation (23) possesses the solution $y(x) \equiv y_0$ on $0 \leqq x < \infty$ satisfying (2) and (3). Considerations similar to those concerning (21) show that no solution of (23) satisfying $y(C) = y_0$ can vanish for any x if $y_0 > 1$.

10. An important question concerning the solutions of (6) which satisfy (3) for $0 \leqq x < \infty$ is whether or not the limit relation

$$\lim_{x \to \infty} y(x) = 0 \tag{24}$$

holds. In his theorem concerning $y'' = f(x, y)$, quoted in § 1, Kneser concluded (24) incorrectly; cf. [17], p. 95. For this case, Mambriani [8], p. 620, gave a sufficient condition for (24); namely, that for every $c > 0$ there should exist a positive constant $C = C_c > 0$ such that

$$f(x, y) \geqq C\sigma(x) \text{ for } c \leqq y < \infty,$$

where $\sigma(x) \geqq 0$ is a continuous function on $0 \leqq x < \infty$ satisfying

$$\int_0^\infty \sigma(x)\,dx = \infty. \tag{26}$$

This condition can be improved greatly; cf. (29) below.

The improvement will be based on the following facts: (i) if $q(x)$ is continuous for large positive x, then the linear differential equation (4) must possess a solution $y(x)$ tending to a finite, non-vanishing limit if

$$\int^\infty x\,|\,q(x)\,|\,dx < \infty;$$

(ii) the latter sufficient condition is necessary as well if $q(x) \geqq 0$ (in other words, if $q(x) \geqq 0$, then

$$\int^\infty xq(x)\,dx < \infty \tag{27}$$

is necessary and sufficient for the existence of a $y(x)$ for which $y(\infty)$ exists and in neither 0 nor ∞).

Assertions (i) and (ii) can be concluded from [1], § 4 and [15], § 1, respectively. A refinement of (i) and a simple proof of (ii) are given in [18].

THEOREM IV. *Let $f(x, y, z)$ be a continuous function on R, with the property that, for every pair of constants $C > c > 0$, there exists a non-negative, continuous function $\sigma(x) = \sigma(x; c, C)$ satisfying*

$$f(x, y, z) \geqq \sigma(x) \text{ for } 0 \leqq x < \infty,\ c \leqq y \leqq C,\ -C \leqq z \leqq 0 \tag{28}$$

and

$$(29) \qquad \int^{\infty} x\sigma(x)\,dx = \infty.$$

Then (24) *holds if* $y = y(x)$ *is any solution of* (6) *on* $0 \leqq x < \infty$ *which satisfies* (3).

Proof of Theorem IV. Since the functions $\sigma(x) = \sigma(x; c, C)$ are non-negative, it follows from (28), (22) and (6) that $y''(x) \geqq 0$. Hence, if $C \geqq -y'(0)$, it is clear that $-C \leqq y'(x) \leqq 0$ for $0 \leqq x < \infty$.

Suppose that the assertion of Theorem IV is false. Then $y(\infty) = c > 0$ holds for some $y(x)$. Hence, if $C \geqq y(0)$, then $c \leqq y(x) \leqq C$ for $0 \leqq x < \infty$. Choose $c = y(\infty) > 0$ and $C = \max(y(0), -y'(0))$.

By (6), the function $y = y(x)$ satisfies the linear differential equation (4), where $q(x) = f(x, y(x), y'(x))/y(x)$. It is seen from (28) that $q(x) \geqq \sigma(x)/y(x) \geqq \sigma(x)/c$. Hence (29) implies that (27) cannot hold. Consequently, $y(\infty) = c > 0$ is impossible. This contradiction proves Theorem IV.

11. If (4) is replaced by (5), then (27) becomes

$$(30) \qquad \int_0^{\infty} \{ \int_0^{x} \exp(-\int_u^{x} r(s)\,ds)\,du \} q(x)\,dx < \infty$$

(which reduces to (27) if $r(x) \equiv 0$). By taking advantage of this fact, one can prove a theorem similar to Theorem IV. In contrast to the linear case, the resulting theorem is not more general than Theorem IV.

THEOREM IV*. *Let* $f(x, y, z)$ *be a continuous function on* R, *with the properties that* (8) *holds and that, for every pair of constants* $C > c > 0$, *there exists a pair of continuous functions,* $\sigma(x) = \sigma(x; c, C)$ *and* $\tau(x) = \tau(x; c, C)$, *satisfying*

$$(31) \qquad f(x, y, 0) \geqq \sigma(x) \geqq 0 \text{ for } 0 \leqq x < \infty,\ c \leqq y \leqq C;$$

and

$$(32) \quad f(x, y, z) - f(x, y, 0) \geqq \tau(x) z \text{ for } 0 \leqq x < \infty,\ c \leqq y \leqq C, \quad -\infty < z < 0;$$

finally,

$$(33) \qquad \int_0^{\infty} \{ \int_0^{x} \exp(-\int_u^{x} \tau(s)\,ds)\,du \} \sigma(x)\,dx = \infty.$$

Then (24) *holds if* $y = y(x)$ *is any solution of* (6) *on* $0 \leqq x < \infty$ *which satisfies* (3).

It is not assumed that $\tau(x) \geqq 0$ in (32). Incidentally, if $f(x, y, z)$ satisfies either (22) or Nagumo's condition (10)-(11), then $y'(x)$ is bounded (cf. Hilfssatz 1, [9], p. 862). In this case, the range $-\infty < z < 0$ in (32) can be replaced by $-C \leqq z < 0$.

Proof of Theorem IV*. Suppose that the assertion of Theorem IV* is false. Then $y(\infty) = c > 0$ for some $y(x)$. This solution $y = y(x)$ of (6) cannot be constant for large values of x. For otherwise $f(x, c, 0) \equiv 0$ for large x, and this contradicts (31) and (33). Since Lemma 2 is applicable, it follows that $y'(x) < 0$ for $0 < x < \infty$.

Since $y = y(x)$ satisfies (6), it satisfies (5) if $q(x)$, $r(x)$ are defined by (17), (18). Then $q(x) \geqq \sigma(x)/y(x) \geqq \sigma(x)/c$, by (8) and (31). But $r(x) \leqq \tau(x)$ by (32). Hence the factor $\{\cdot\ \cdot\ \cdot\}$ in (30) is not less than that in (33). Thus (33) shows that (30) cannot hold. Since this implies that $y(\infty) = 0$, Theorem IV* follows.

12. By combining Theorem III with a procedure applied in [17] and [5] to the case of a linear differential equation, the following theorem will now be proved:

THEOREM V. *Let* $g(x, y, z)$ *be defined and of class* C^∞ *on* R, *and suppose that*

$$g(x, 0, 0) \equiv 0 \text{ for } 0 \leqq x < \infty \tag{34}$$

and that the partial derivatives of g *satisfy the inequalities*

$$(-1)^{i+k}\partial^{i+j+k}g/\partial x^i \partial y^j \partial z^k \geqq 0 \text{ for } i, j, k = 0, 1, \cdots \text{ on } R. \tag{35}$$

Let $p(x) > 0$ *be a continuous function on* $0 \leqq x < \infty$, *and suppose that* $p(x)$ *possesses a completely monotone first derivative* $p'(x)$ *on* $0 < x < \infty$, *that is,*

$$(-1)^{n-1}d^n p/dx^n \geqq 0 \text{ for } n = 1, 2, \cdots \tag{36}$$

($0 < x < \infty$). *Then there exists a positive* $Y (\leqq \infty)$ *such that, if* $0 < y_0 < Y$, *the differential equation*

$$p(x)y'' = g(x, y, y') \tag{37}$$

possesses a solution $y = y(x)$ *satisfying* (2) *and representable in the form*

$$(38) \qquad y(x) = \int_0^\infty e^{-xt} d\psi(t) \text{ for } 0 < x \leqq \infty$$

where $\psi(t)$ *is non-decreasing for* $0 \leqq t < \infty$.

Proof of Theorem V. First, $f(x, y, z) = g(x, y, z)/p(x)$ satisfies the conditions of Theorem III. In fact, (8) follows from (34), and (22) from the case $i = 0$, $j = 0$, $k = 0$ of (35). Clearly, the definition of $f(x, y, z)$ can be extended, to a domain which contains R in its interior, in such a way that f_y and f_z continue to exist and to be continuous.

Thus Theorem III is applicable. Hence there exists a positive $Y (\leqq \infty)$ with the property that (37) possesses, if $0 < y_0 < Y$, a solution satisfying (2) and (3). It is clear that $y(x)$ is of class C^∞ for $0 < x < \infty$.

In order to prove that $y(x)$ has a representation of the form (38), it is sufficient to show that $y(x)$ is completely monotone for $0 < x < \infty$:

$$(39) \qquad (-1)^n d^n y/dx^n \geqq 0 \text{ for } n = 1, \cdots \text{ if } 0 < x < \infty$$

(Hausdorff-Bernstein). The inequalities (3) are the cases $n = 0, 1$ of (39). The cases $n \geqq 2$ follow by induction from (3), (35) and (36), by successive differentiations of (37); cf. the arguments in [6], pp. 350-351. This proves Theorem V.

Remark. The function $f(x, y, z)$ occurring in the preceding proof satisfies the conditions of Theorem I (which assures that $Y = \infty$) if and only if

$$(40) \qquad \partial^3 g(0, y, z)/\partial z^3 \equiv 0 \text{ for } 0 < y < \infty, -\infty < z \leqq 0.$$

In order to see this, note that (34) and (35) imply that (8) and (9) hold, and that

$$0 \leqq f(x, y, z) \leqq f(0, C, z) \text{ for } 0 \leqq x \leqq C,\ 0 \leqq y \leqq C,\ -\infty < z \leqq 0.$$

Furthermore, $f(0, C, z) = g(0, C, z)/p(0)$ is a non-increasing function of z for $-\infty < z \leqq 0$. Hence Nagumo's condition (10)-(11) can be satisfied if and only if

$$(41) \qquad \int_{-\infty} z dz/g(0, C, z) = -\infty \text{ for every } C > 0.$$

The inequalities (35) imply that

$$(42) \quad (-1)^k \partial^k g(0, C, z)/\partial z^k \geqq 0 \text{ for } k = 0, 1, \cdots \text{ if } -\infty < z \leqq 0.$$

Hence, unless

$$\partial^3 g(0, C, z)/\partial z^3 \equiv 0 \text{ for } -\infty < z \leqq 0, \tag{43}$$

it follows that

$$g(0, C, z) \geqq \text{const. } |z|^3, \text{ where const.} > 0. \tag{44}$$

This implies that (41) cannot hold.

In order to see that the negation of (43) implies (44), suppose that $\partial^3 g(0, C, z_0)/\partial z^3 \leqq \text{Const.} < 0$ for some z, say $z = z_0 \leqq 0$. Then

$$\partial^3 g(0, C, z)/\partial z^3 \leqq \text{Const.} < 0 \text{ for } -\infty < z \leqq z_0,$$

by the case $k = 4$ of (42). This proves (44). Hence, (41) cannot hold. On the other hand, it is clear that (43) is sufficient for (41). This proves the Remark concerning (40).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Bôcher, "On regular singular points of linear differential equations of the second order whose coefficients are not necessarily analytic," *Transactions of the American Mathematical Society*, vol. 1 (1900), pp. 40-52.

[2] R. Caccioppoli, "Problemi non-lineari in analisi funzionale," *Rendiconti del Seminario della R. Università di Roma*, ser. 3, vol. 1 (1933), pp. 13-22.

[3] S. Cinquini, "Sopra i problemi di valori al contorno per equazioni differenziali del secondo ordine," *Annalo della R. Scuola Normale Superiore di Pisa*, ser. 2, vol. 8 (1939), pp. 271-283.

[4] P. Hartman and A. Wintner, "On the asymptotic behavior of the solutions of a non-linear differential equation," *American Journal of Mathematics*, vol. 68 (1946), pp. 301-308.

[5] ——— and A. Wintner, "On the Laplace-Fourier transcendents," *ibid.*, vol. 71 (1949), pp. 367-372.

[6] ——— and A. Wintner, "On non-linear differential equations of first order," *ibid.*, vol. 72 (1950), pp. 347-358.

[7] A. Kneser, "Untersuchung und asymptotische Darstellung der Integrale gewisser Differentialgleichungen bei grossen reellen Werthen des Arguments, I," *Journal für die reine und angewandte Mathematik*, vol. 116 (1896), pp. 178-212.

[8] A. Mambriani, "Su un teorema relativo alle equazioni differenziali ordinarie del 2° ordine," *Atti della Reale Accademia Nazionale dei Lincei*, ser. 6, vol. 9 (1929), pp. 620-622.

[9] M. Nagumo, "Ueber die Differentialgleichung $y'' = f(x, y, y')$," *Proceedings of the Physico-Mathematical Society of Japan*, ser. 3, vol. 19 (1937), pp. 861-866.

[10] A. Rosenblatt, "Sur les théorèmes de M. Picard dans la théorie des problèmes aux limites des équations différentielles ordinaires non linéaires," *Bulletin des Sciences Mathématiques*, ser. 2, vol. 57 (1933), pp. 100-106.

[11] G. Scorza-Dragoni, "Su un'equazione differenziale particolare," *Atti della Reale Accademia Nazionale dei Lincei*, ser. 6, vol. 9 (1929), pp. 623-625.

[12] ———, "Il problema dei valori ai limiti studiato in grande per gli integrali di una equazione differenziale del secondo ordine," *Giornale di Mathematiche*, ser. 3, vol. 69 (1931), pp. 77-112.

[13] ———, "A propositio di alcuni teormi relativi ad un problema ai limiti per una equazione differenziale del secondo ordine," *Atti della Reale Accademia Nazionale dei Lincei*, ser. 6, vol. 22 (1935), pp. 44-48.

[14] ———, "Elementi uniti di trasformazioni funzionali e problemi di valori ai limiti," *Rendiconti del Seminario Matematico della R. Università di Roma*, ser. 4, vol. 2 (1938), pp. 255-275.

[15] H. Weyl, "Ueber gewöhnliche lineare Differentialgleichungen mit singulären Stellen und ihre Eigenfunktionen," *Göttinger Nachrichten*, 1909, pp. 37-63.

[16] A. Wintner, "The infinities in the non-local existence problem of ordinary differential equations," *American Journal of Mathematics*, vol. 68 (1946), pp. 173-178.

[17] ———, "On the Laplace-Fourier transcendents occurring in mathematical physics," *ibid.*, vol. 69 (1947), pp. 87-98.

[18] ———, "On almost free linear motions," *ibid.*, vol. 71 (1949), pp. 595-602.

[19] G. Zwirner, "Un criterio di esistenza per un problema di valori al contorno per equazioni differenziali del secondo ordine," *Rendiconti del Seminario Matematico della R. Università di Padova*, vol. 10 (1939), pp. 55-64.

THE TOPOLOGY OF REGULAR CURVE FAMILIES WITH MULTIPLE SADDLE POINTS.*[1]

By WILLIAM M. BOOTHBY.

Introduction. It is known that the level curves of any function $f(x, y)$ which is harmonic in a simply connected domain form a curve family which is regular (locally homeomorphic to parallel lines) in the neighborhood of every point of the domain with the exception at most of a set of isolated points at each of which the curve family has a singularity of the multiple saddle point type (Figure 1). The central task of this and a later paper[2] is to prove that these local properties are sufficient to characterize topologically the level curves of such harmonic functions, and to develop certain general topological properties of these families of curves which should prove useful in the study of harmonic functions and analytic functions. These investi-

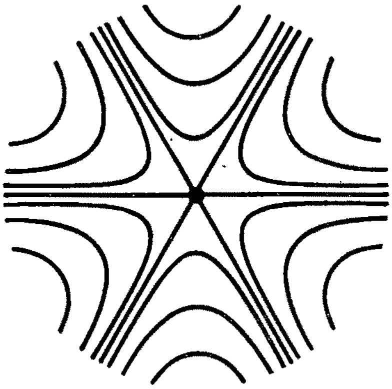


FIGURE 1
MULTIPLE SADDLE POINT

---

* Received January 30, 1950.

[1] The material in this paper and the one to follow was taken from the author's Ph. D. thesis at the University of Michigan. The author wishes to express his gratitude to Professor Wilfred Kaplan for his guidance in this research and his advice in the preparation of this paper.

[2] The Topology of the Level Curves of Harmonic Functions with Critical Points, to appear in this JOURNAL.

gations generalize some of the results of several papers of W. Kaplan in which curve families which were regular (without singularities) throughout a simply connected domain were considered. In particular it was proved by Kaplan that (1) every such curve family is homeomorphic to the level curves of a function harmonic in a simply connected domain [VIII]; (2) every such family is homeomorphic to the solution family of a system of differential equations $dx/dt = f(x, y)$, $dy/dt = g(x, y)$ [VII]; and (3) that every such family can be decomposed into the sum of a denumerable collection of non-overlapping subfamilies each homeomorphic to the parallel lines of a half-plane. These results are all extended in this and the following paper to the more general type of curve family with isolated singularities of the multiple saddle point type. More precisely, the generalization of (1), (2), and (3) is achieved in the following paper; in this paper the necessary preliminary theorems are proved about the general topological structure of such families, the principal theorem being that the curve family can always be considered as the family of level curves of a continuous function without relative extrema.

Section 1 is devoted to definitions and to enumeration without proof of certain important properties of curve families in general. The most important of these deal with the sets of limit points $L(C+)$, $L(C-)$, $L(C)$, of the individual (directed) curves C of the family. These theorems stem essentially from the work of Poincaré and Bendixson. They depend, however, on the regularity only, as noted by Kaplan.

In Section **2** we restrict our curve families to have isolated singularities only and note that this implies that the set of singularities must then be closed and denumerable. The important notion of index (relative to the curve family) of a closed curve and of a singularity is defined and the theorems we will need stated, in particular the fact that the index of a closed curve is equal to the sum of the indices of the singularities in its interior. This notion of index was defined for curve families by Hamburger [see IX] and it coincides with the Poincaré index when the curve family is the family of integral curves of a vector field. It was studied in detail by Y. S. Chin [II] from which source the theorem mentioned was taken.

Finally in Section **3** we restrict our curve families to be regular in a simply connected domain except for multiple saddle points, which, for reasons explained below, will henceforth be called branch points. Such a family will be called a branched regular curve family and is the principal object of interest in what follows. The singularities will by their very nature be isolated, hence form a closed denumerable set. The index forms a useful tool for ruling out many undesirable configurations of curves, such as closed curves, "polygons" of curves (with singular points as vertices), and so on. Moreover it enables us to investigate some of the relations between the curve family and cross-sectional arcs (i. e. arcs which locally intersect each curve only once). Finally it enables us to show that each curve either extends to infinity in a given direction on it or ends at a branch point.

From this it follows easily that the curves are of two types, one the regular curves which extend to infinity in each direction, the other branched curves which end in a branch point at one or both ends. These latter then naturally form configurations which are trees (without ends) in the topological sense, i. e. they are locally finite, one dimensional simplicial complexes, with curves of the family as 1-cells, branch points as 0-cells. It is shown by use of a theorem of Adkisson and MacLane [I] that any such tree may be straightened out to a rectilinear one by a homeomorphism of the entire plane.

Finally in Section 4 it is shown that the trees are distributed in such a fashion on the plane that, if it is done with sufficient care, we may remove certain curves of each tree with their endpoints, i. e. make cuts in the plane extending to infinity along the tree, in such a fashion that the plane with these curves removed is simply connected and the remaining curves of the family form a regular curve family filling this simply connected subset of the plane. It follows from Kaplan [IV] that these curves are the level curves of a continuous function without relative extrema, and by extending this function to the entire plane, we obtain our family too as the level curves of a continuous function without relative extrema.

1. General Properties of Regular Curve Families in the Plane.

1. 1. Curve families filling a region. An *open curve* will mean a homeomorphic image of an open interval, a *closed curve* a homeomorphic image of a circle, a *half-open curve* a homeomorphic image of a half-open interval, and an *arc* of a closed interval. A *curve* will mean any one of these four. A family F of curves will be said to *fill* a subset R of the Euclidean plane π if every curve of F is in R and every point of R lies on one and only one curve of F. If U is a subset of π such that each curve C of a family F filling R intersects U in a set $U \cap C$ each of whose components is a curve, then we denote by $F[U]$ the curve family filling $U \cap R$ whose curves are the components of $C \cap U$ for all C in F. If the curve family F fills R and the curve family G fills S, then F and G will be called *homeomorphic* if there is a homeomorphism of R onto S such that the image of each curve in F is a curve in G. If p is a point of R, R filled by a curve family F, then C_p will denote the curve of F through p.

1.2. Regularity. Henceforth, F will denote a curve family filling a subset R of π, the *oriented* Euclidean plane. We shall use R_0 to denote the rectangle $|x| \leqq 1$, $|y| \leqq 1$, of the xy-plane, and F_0 the family of lines $y = \text{constant}$ filling R_0. F will be said to be *regular* at a point p of R if there is a set $U(p)$ in R to which p is interior (relative to R) and such that $F[U(p)]$ is homeomorphic to F_0. F is then *regular in* R if F is regular at every point of R. A *cross-section* of F (*through* the point r) is an arc pq (to which r is interior), such that pq lies in a subset R' of R which is open relative to R, and such that each curve of $F[R']$ meets pq at most once. An *r-neighborhood* of a point p of R will mean a subset $U(p)$, $\overline{U(p)} \subset R$, which contains p and is open relative to R, and is moreover such that $F[\overline{U(p)}]$ is homeomorphic to the family F_0, filling the rectangle R_0 (above) of the xy-plane, in such a way that the inverse images of the lines $|x| = 1$ are cross-sections. The following properties of regular curve families will be useful later.

THEOREM 1.2-1. *If a family F fills an open region R and is regular in R, then each curve of F is either open or closed in π.* [IV, 1]

THEOREM 1.2-2. *If a family F fills any region R and is regular in R, then every point p of R has an arbitrarily small r-neighborhood $U(p)$, and there is a cross-section pq with p as endpoint. If p is in the interior of R, then there is a cross-section through p. Moreover if st is any arc lying on a curve C of F, then there is, within any ϵ-neighborhood $U_\epsilon(st)$, an r-neighborhood containing st.* [IV, 8 and VI, 1]

THEOREM 1.2-3. *Let F be a curve family filling R where $R = R_1 \cup R_2$ and $F[R_1]$ and $F[R_2]$ are both defined. If p is an interior point of R, and $F[R_1]$ and $F[R_2]$ are both regular at p, then F is regular at p.* [VI, 2]

1.3. The sets $L(C+)$ and $L(C-)$. From this point on we assume every curve of F directed, although we may later find it at times convenient to reverse directions on a given curve. If C is any open curve in F, then by a *positive* (*negative*) *limit point* of C will be meant any point q which is the limit of a sequence $p_n = f(t_n)$, where C is the image under f of $0 < t < 1$ with increasing t corresponding to the positive direction on C and $t_n \rightarrow 1$ ($t_n \rightarrow 0$). The set of all positive (negative) limit points of the directed curve C will be denoted by $L(C+)$ (by $L(C-)$). $L(C)$ is defined by $L(C) = L(C+) \cup L(C-)$ and $C \cup L(C)$ is the point set closure of C. Clearly, $L(C) \cap C$ is empty since C is homeomorphic to $0 < t < 1$. The theorems below involving $L(C+)$ hold equally well for $L(C-)$.

THEOREM 1. 3-1. *If C is an element of a regular curve family F, and $L(C+)$ contains a closed curve D of F, then $L(C+) = D$.* [IV]

THEOREM 1. 3-2. *If F is a regular curve family filling R and p is in R and, moreover, in $L(C+)$ for some curve C of F, then every point of the curve D_p of F through p is in $L(C+)$.* [IV, 7]

2. Isolated Singularities and Index.

2. 1. Isolated singularities. Let F be a curve family regular in R and b be an isolated boundary point of R. Then, if for some neighborhood $U(b)$ in π, F is regular in $U(b) - b$ but not in $U(b)$, we shall call b an *isolated singularity* of F. In [VI] W. Kaplan has completely classified the structure of a regular curve family in the neighborhood of an isolated singularity. Below we shall actually be interested in a very restricted type of isolated singularity; however, in this section no restrictions will be made.

THEOREM 2. 1-1. *Let F be a regular curve family filling the region R, where R consists of the entire plane minus a set of points B, each of which is an isolated singularity of F. Then B is a closed and denumerable point set.*

Proof. From the fact that each point of B has a neighborhood containing no other point of B it follows at once that B is denumerable and that no point of B is a limit point of B. Now consider a point p at which F is regular. We must show that p is not in $\overline{B}$, whence $\overline{B} = B$. By the definition of regularity there is a set $U \subset R$ such that p is interior to U relative to R and there is a homeomorphism $f\colon U \rightarrow R_0 = \{(x,y) \mid |x| \leqq 1; |y| \leqq 1\}$. It is clear that U is a Jordan domain bounded by the inverse image of the closed curve which forms the boundary of R_0. Moreover we can, as is well-known, extend f to map all of π on the xy-plane. Hence, for simplicity, we shall deal with the homeomorphic image of π, R, F, etc. on the xy-plane. If $p' = f(p)$ is interior to $R_0 = f(U)$, then p' is not in $\overline{B'} = \overline{f(B)} = f(\overline{B})$ since the complete description of U, the r-neighborhood of p, makes it clear that the interior of U (in π) is in R. Now assume p' lies on the boundary of R_0. Since p' is by definition of U, an interior point of R_0 relative to R', there is a small circle which with its interior intersects R' in a set contained in R_0. Then clearly, if this circle is of radius less than one, at least two quadrants of it lie outside R_0 hence outside R', i. e. in B'. It follows that B' cannot be a set of isolated points. Therefore our assumption that there was such a point on the boundary of R_0 is false, and p' is interior to R_0 and not in $\bar{B}'$. Hence $B' = B'$ and $B = B$, which completes the proof.

2. 2. Index. Following the definition given in Kerékjártó [IX, p. 251 ff.], we define the *index* of an isolated singularity b of a family F as follows: Let K be any simple closed curve containing the isolated singularity but no other singularities on it or in its interior and let $U_1, \cdots, U_n$ be a covering of K by r-neighborhoods. Then, if the neighborhoods have been chosen sufficiently small, it is clear that we may replace K by a simple closed curve K' in $\bigcup_{i=1}^{n} U_i$ and such that (i) K' is a "polygon" composed of sides which are alternately (1) arcs of curves of F and (2) cross-sections of F, and (ii) K, K' each contain in their interiors only the singularity b. Every vertex of the polygon K' is the intersection of a cross-section and a curve of F; we call it an *internal* vertex, if the curve of F which forms the side at that vertex enters the interior of K' at the vertex, and in the other case we call it an *external* vertex. If we denote the number of internal vertices by e and external vertices by a, then let $\rho(K') = 1 - (a - e)/4$, it is shown by Kerékjártó that $\rho(K')$ so defined is independent of the particular K' chosen, depending only on the singularity b. We may then denote this number by $\rho(b)$ and define it to be the index of b.

More generally, given any closed curve K in the region R filled by the regular curve family F, we may in the same fashion define a curve K' which is close to K, which contains in its interior exactly those singularities contained interior to K, and which consists itself of alternate arcs and cross-sections. Exactly as above we may define $\rho(K') = 1 - (a - e)/4$ and show that this quantity depends only on K, not on the choice of K'. Hence we denote it $\rho(K)$ and call it the index of the curve K. Since we have restricted ourselves to families with isolated singularities, there will of course be only a finite number of singularities interior to any closed curve. The index of a closed curve relative to a given curve family has been studied in detail by Y. S. Chin [II] and he proves the following:

THEOREM 2. 2-1. *Let K be a simple closed curve in R, then $\rho(K) = \sum_{i=1}^{n} \rho(b_i)$ where b_i, $i = 1, \cdots, n$ are the singularities of F interior to K.*

This theorem will not be proved since a detailed proof may be found in [II]. We also state the following useful theorem from Kerékjártó.

THEOREM 2. 2-2. *Let b be an isolated singularity of the regular curve family F and let b be the positive endpoint of exactly p curves of F and the negative endpoint of exactly q curves of F. Then* $\rho(b) = 1 - (p+q)/2$.

3. Regular Curve Families Whose Singularities Are Branch Points.

3. 1. Branch points. Let b be a boundary point of $R \subset \pi$ and F a curve family regular in R. If there exists a neighborhood $U(b)$ such that $F[U(b) - b]$ is homeomorphic to the level curves of $\Re(z^n)$, the real part of $f(z) = z^n$, $n > 1$, under a homeomorphism g carrying $U(b)$ onto $|z| < 1$ with $g(b) = 0$, then we say that b is a *branch point* of F with multiplicity n. This configuration is probably better known as a multiple saddle point, but the term branch point will be adopted below for two reasons: first, these points appear as branch points on trees and, second, they will have branch points of the inverse function as images if we are concerned with the level curves of the real part of an analytic function. The term saddle point was used in the title because it seems to be more suggestive of the type of curve family considered. (See Figure 1.)

The neighborhood $U(b)$ above will be called an *admissible* neighborhood; more precisely $U(b)$ plus the homeomorphism g, which shall always be taken as an orientation preserving homeomorphism, will constitute an admissible neighborhood. In the case of a branch point b of multiplicity n, there are exactly $2n$ curves of $F[U(b)]$ which end at b, i. e. which may be directed so that $L(C+) = b$. It follows that the multiplicity is independent of the choice of $U(b)$. It is clear that b is an isolated singularity of F.

From now on we shall consider only curve families F which are regular in a region R which in general will contain all of π except for a set B of branch points of F. At times it may be expedient to let $R \cup B$ be not all of π but a simply connected domain in π. Of course, topologically the two are equivalent and for convenience we shall consider almost always the first case. Such families will be called *branched regular curve families* filling[3] π (or an open, simply-connected subset of π). The following theorem is well-known:

THEOREM 3. 1-1. *The family F of level curves of a function $f(x, y)$ harmonic in a simply connected*[4] *domain D is a branched regular curve family filling D. The critical points of $f(x, y)$ are the branch points of F.*

---

[3] Strictly speaking, according to the definition in 1. 1, the curve family fills not π but *π minus the branch points*. For convenience the latter phrase will often be dropped below.

[4] Simple connectedness is not needed for this theorem, but is needed for the converse.

It is this theorem which motivates in large part this study of such curve families. A converse to this theorem will be proved in a subsequent paper.

3. 2. General properties of branched regular curve families. Restricting ourselves now to branched regular curve families F filling π, i. e. more precisely $R = \pi - B$, B the set of branch points, we make a number of preliminary definitions. The curves of F are assumed directed, so that there shall be no ambiguity in the use of the symbols $L(C+)$ and $L(C-)$. We note that since R is open by Theorem 2. 1-1, every curve of F is by Theorem 1. 2-1 either open or closed. If $L(C) = 0$ we call C a regular curve, and if $L(C+) = b \varepsilon B$, i. e., a branch point, we shall say that C is a *branched curve,* branched at the positive end at b; we also call b the *positive endpoint* of C in this case. Similarly if $L(C-) = b' \varepsilon B$. It will subsequently be shown that the only possibility is that $L(C+) = 0$ or $= b$, a single branch point, and similarly $L(C-)$, so we shall not give any name to the as yet possible type of curve which might have more than one positive (or negative) limit point).

If $b \varepsilon B$, the curves of F which have b as endpoint, together with their endpoints, form a set called the *star of* b, $St(b)$; and the same curves without their endpoints, except b, form the *open star of* b. If b is of multiplicity n, then there are at most $2n$ curves in $St(b)$; it will be shown later in this section that there are exactly $2n$, i. e., the two endpoints of a curve cannot coincide. It is useful to note that since $St(b)$ contains a finite number of curves, plus the fact that B is denumerable, there are at most a denumerable number of branched curves in F.

If $C_1, \cdots, C_n$ are $n \geqq 2$ distinct branched curves of F with their endpoints, and may be so directed that $L(C_i+) = b_i = L(C_{i+1}-)$, $b_i \varepsilon B$ and b_i distinct for $i = 1, \cdots, n-1$, and if in addition $L(C_1-) \neq b_i \neq L(C_n+)$, $i = 1, \cdots, n-1$, then we call $C_1, \cdots, C_n$ either a *simple polygon* of branched curves or a *chain* of branched curves according to whether or not there is a $b_0 \varepsilon B$ such that $L(C_1-) = b_0 = L(C_n+)$. In brief, the curves $C_1, \cdots, C_n$, together with their endpoints will form a chain if the collection is homeomorphic to an arc, and a simple polygon if homeomorphic to a circle. For convenience we shall call a single curve a polygon if it has coincident endpoints. The theorem which follows shows that the polygon configuration cannot occur in F.

THEOREM 3. 2-1. *A branched regular curve family F filling π can contain neither a closed curve nor a simple polygon of branched curves.*

Proof.[5] .Suppose F contains a closed curve C. Kaplan has shown [IV] that if a regular curve family contains a closed curve, the interior of the curve must contain a boundary point, i. e., in this case an element of B. Thus it follows from Theorem 2. 2-2 that the sum of the indices of the branch points interior to the closed curve is a negative integer, since by that theorem the index of a branch point of multiplicity n is $1-n$. However, the index of the closed curve itself is one, which is a contradiction to Theorem 2. 2-1. Thus F contains no closed curve.

Next we suppose that F contains a polygon K of branched curves. We construct a curve family filling the circular region $R=\{(x,y)\mid x^2+y^2\leqq 4\}$ of the xy-plane as follows: fill the annular region $R_1=\{(x,y)\mid 1<x^2+y^2\leqq 4\}$ with circles $x^2+y^2=c^2$, $1<c\leqq 2$, then map K with its interior onto R_2, the unit circular disc, K mapping onto the circumference. Then $R=R_1\cup R_2$, except for the images of the singular points interior to K plus certain of the images of those on K, is filled by a curve family F'. $F'[R_1-S]$ and $F'[R_2]$ (where $S=$ singular points mentioned above) are regular and it follows from Theorem 1. 2-3 that F' is regular in $R-S$. We now examine the singularities more closely. Let b be a singular point on K, i. e., the common endpoint of two curves of the polygon; b is a branch point of multiplicity n and thus $St(b)$ contains more than the two curves of K which meet at b. If $St(b)\cap \operatorname{int} K=0$ it is clear that the image b' of b will not be a singular point of F'; however, if this is not the case, b' will be, and there will be at least three curves of F' which have the image b' of b as endpoint. Thus by Theorem 2. 2-2 the index $\rho(b')$ of this singularity of F' is $\leqq -\frac{1}{2}$. Hence if we consider the curve $x^2+y^2=4$, whose index is one, we again reach a contradiction, for this curve contains singularities, all of whose indices are negative and thus the requirements of Theorem 2. 2-1 cannot be satisfied. Should the image K' of K contain no singularities of F', we are again reduced to the case proved earlier, since F' will in fact be a branched regular curve family. Whence the theorem is proved.

We shall now describe some other simple configurations which cannot occur in a branched regular curve family. Let us suppose that a cross-section pq crosses a curve C of F at exactly two points t and u between p and q. Denote by $(tu)_1$ the arc of pq between t and u, and by $(tu)_2$ the arc on C. Then $K=(tu)_1\cup(tu)_2$ is a simple closed curve. We distinguish three possible cases: (i) if p and q are exterior to K, we call this pattern a *bay*, (ii) if one of the points is interior and the other exterior, we call it a *spiral*, and (iii) if both p and q are interior to K, we call it a *harbor* (see Figure 2). From the definition it is seen that the index $\rho(K)$ is $\frac{1}{2}$, 1 and $1\frac{1}{2}$ respectively in the three cases. Since the sum of the indices of the branch points interior to K must be 0 (if there are none) or a negative integer, we have at once:

---

[5] It is suggested that many of the proofs can be followed much more easily by the reader if he makes his own diagrams where none is provided.

THEOREM 3.2-2. *A branched regular curve family filling π can contain neither a bay, nor a spiral, nor a harbor.*

THEOREM 3.2-3. *The set $L(C+)$ is either empty or contains a single branch point. Similarly $L(C-)$.*

Proof. Suppose $L(C+)$ contained a branch point b and in addition contained points other than b. Let p_n be a sequence of points on C which approach b, which we assume to have multiplicity n. We consider an admissible neighborhood $U(b)$ with its homeomorphism f carrying U onto the interior of the unit circle. Let F' denote the level curves of $\Re(z^n)$ onto which $F[U]$ is carried homeomorphically by f. Since $L(C+)$ contains points other than b it is clear that $f(C)$ is not one of the radial curves, $\theta=(k/n)\pi$, in F' but must have its image in one or more of the sectors formed by these radial curves, and in fact since $f(p_n)\rightarrow f(b)$ the origin, $f(C)$ must cross one of the sectors an infinite number of times. Hence it must be possible to choose points q_n on C near p_n so that q_n approach a regular point q whose image lies on one of the radial lines. Thus we are forced under our assumption to conclude that $L(C+)$ contains at least one regular point q in $F[U(b)]$. Finally, it is clear from the structure of $f(U)$ that this implies that a cross-section through q crosses C an infinite number of times and hence must form at least one bay, spiral, or harbor in so doing. This contradicts the previous theorem; whence it follows that if b is in $L(C+)$, it is the only point in this set.

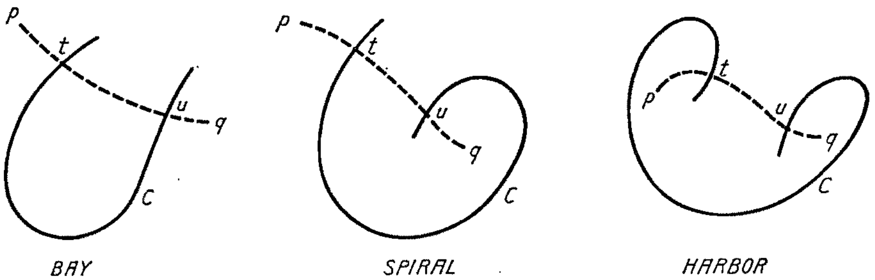


FIGURE 2

We remark that this proves in particular that if a curve C is bounded in the positive direction, then $L(C+)$ must be a branch point.

THEOREM 3.2-4. *An arc pq of $R = \pi - B$ is a cross-section if and only if it meets each chain (or curve) at most once.*

Proof. If pq satisfies the conditions, it is a cross-section by definition. Now suppose pq is a cross-section and that it crosses some chain twice. Let t and u be successive points of crossing and denote by $(tu)_1$ the arc from t to u on pq, and by $(tu)_2$ the arc on the chain. We then have a simple closed curve $K = (tu)_1 \cup (tu)_2$. Interior to K there are only a finite number of branch points, hence by virtue of the preceding theorem we can surely find a curve C through some point r interior to K which leaves K in both directions. Letting r_1 be the first point in the positive direction from r at which C intersects $(tu)_1$, and r_2 the first in the negative direction, then the arcs from r_1 to r_2 on $(tu)_1$ and C must form a bay, a spiral, or a harbor. Since this is a contradiction to Theorem 3.2-2 the theorem follows.

3.3. Trees. In this section we define an equivalence relation which decomposes the oriented plane, π, into a collection of disjoint closed sets, each of which is a sum of curves of F and points of B, and each of which is a topological tree of a certain type which we define below:

Definition. Let the closed set T of the oriented plane be decomposable into the sum of an at most denumerable collection of subsets C_i, each closed in π, and satisfying the following four conditions:

(1) Each set C_i is the homeomorphic image of either a closed, half-open, or open line segment (whence we will refer to it as a curve).

(2) Each set C_i has at most an endpoint in common with any C_j, $i \neq j$; and if we denote by $St(b)$ the collection of all curves with b as endpoint, then $St(b)$ consists of a finite even number of curves $\geqq 4$.

(3) There is a unique finite chain $c(C_i, C_j) = (C_i, C_{i_2}, \cdots, C_{i_k}, C_j)$ from C_i to C_j for every i, j; i.e., each curve of the chain having an endpoint in common with the preceding curve as in the definitions of **3.2**.

(4) The sets *open* $St(b)$, consisting of the curves of $St(b)$ without their endpoints opposite b, and *open* C_i, consisting of C_i without its endpoints, are both open sets in T (as a subspace of π).

Then we say that T is a *tree.* (See Figure 3.) Our use of this term is much less general than is usual, but since we consider only this specialized type of tree throughout, there should be no confusion in the use of the term.

The decomposition of any tree T into sets C_i is *unique* quite clearly, except for the numbering, and therefore we may speak without ambiguity of *the* curves of T and *the* endpoints of curves (or, i. e., branch points) of T. Note that a tree is connected and, in fact, arcwise connected by (3) and that by the uniqueness of the chains of (3) there can be no closed curve in T. Condition (4) plus the fact that T is closed in π is equivalent to the following statement: If (p_n) is any sequence of points of T and $p_n \rightarrow p \varepsilon \pi$, then $p \varepsilon T$ and all the points of p_n after some N will lie either on $St(p)$ or on a single curve C_i of T depending on whether p is or is not an endpoint of some curve of T. In the language of combinatorial topology each tree, as described above, is a locally-finite, connected, one-dimensional complex containing no one-cycles. In order to exhibit this, it would be necessary to introduce arbitrarily an infinite number of vertices tending to infinity on each curve of the tree which is homeomorphic to a half-open line segment. Once this is done, the statement is clearly true.

It is clear that in particular any regular curve C of a curve family F is a tree with C being itself the only curve of the decomposition. Now, among the elements of our family F we define the relation *joins* as follows: C is said to *join* C' if and only if there is a finite chain $c(C, C')$ of curves of F from C to C'. If we add to this definition that every curve joins itself, then this is easily shown to be an equivalence relation on the curves of F. We denote by T_C the equivalence class of C, including with each curve its endpoint, i. e., T_C is the set of all curves of F, together with their endpoints, which join C. These equivalence classes are disjoint sets and will be shown below to be trees in the sense of our definition. Thus it will be shown that the curves of a branched regular curve family F fall into classes, each of which is a tree.

THEOREM 3. 3-1. *An arc pq on π is a cross-section of F if and only if it lies entirely in $R = \pi - B$ and has at most one point of intersection with each set T_C.*

Proof. If pq has at most one point in common with each set T_C, since T_C is itself a sum of curves of F with their endpoints, then it will have at most one point in common with each curve of F and hence be a cross-section by definition.

On the other hand, by Theorem 3. 2-4, it is necessary that pq meet any set T_C at most once if it is a cross-section, since if pq met T_C at points r, s, then either C_r, C_s are the same curve or else there is a chain $c(C_r, C_s)$ either of which is impossible by that theorem.

THEOREM 3. 3-2. *Each set T_C of a branched regular curve family F is a tree in the sense of our definition.*

Proof. In the event that C is a regular curve the theorem is trivial since $T_C = C$, as already noted. Now let T_C contain a singular curve, then it follows that it contains only such and at most a countable number, since there are at most a countable number of singular curves in F. Each curve of F in T_C, together with its endpoint(s), will constitute a curve C_i of the decomposition of T_C. Each such set is closed in π, since we include endpoints, and is homeomorphic to either a closed or half-open segment, the latter if the curve extends to infinity in one direction. Thus (1) is satisfied. Condition (2) is, however, also satisfied since each set C_i, being the point set closure of a curve of F, has at most an endpoint in common with any set C_j, $i \neq j$, and, if b is any endpoint, then $St(b)$ contains at least four curves and always an even number, $2n =$ twice the multiplicity of b as a branch point. Likewise (3) is satisfied, i. e., the existence of a chain $c(C, C')$ from $C \subset T_C$ to $C' \subset T_C$ is part of the definition of T_C; and the uniqueness is due to the fact that there can be no polygons of branched curves of F by Theorem 3. 2-1. Finally, we prove simultaneously that condition (4) is satisfied and that T_C is closed as a subset of π. Let p_n be any sequence of points of T_C with a point p of π as limit point. Now if p is a regular point of F, then we take an r-neighborhood $U(p)$ and note that unless every p_n lies on the same curve of F, which is necessarily C_p itself, then we have more than one curve of T_C intersecting $U(p)$ and hence a cross-section through p which must cross T_C more than once, contrary to Theorem 3. 3-1. If p is a branch point, then taking an admissible neighborhood of p, we observe that unless we assume all the points p_n, $n > N$, to lie on $St(p)$, we arrive at the same contradictory conclusion by considering a cross-section from p into one of the sectors of the admissible neighborhood. Thus we conclude that the theorem must be true.

We return to a discussion of a tree T, in the sense of our definition, to study its general structure and its relation to its complementary domain in π. Whether it is or is not a set T_C, i. e. composed of curves of a branched regular curve family or merely abstractly given, is at the moment immaterial. As previously noted, the decomposition of T into curves is unique, and hence we may refer without ambiguity to *the* curves and *the* branch points (or endpoints) of T. Since T is assumed to be imbedded in an oriented plane, a cyclic order is induced on the curves of $St(b)$ [see I]; this allows us to make the following definitions.

We shall call curves C, C' *clockwise adjacent* if they have a common endpoint b and, in the order induced on $St(b)$ by the orientation of π, moving clockwise the curve C' follows C. In the case of curves of F this means that in each (orientation preserving) homeomorphism of $F[U]$, U an admissible neighborhood of b, onto the level curves of $\Re(z^n)$ in $|z| < 1$, the images of $C \cap U$ and $C' \cap U$ will correspond to radial lines $\theta = (k/n)\pi$, $\theta' = ((k-1)/n)\pi$ respectively. Clearly, given a curve C with endpoint b, there is exactly one curve C' such that C, C' form a clockwise adjacent pair. We call C', C *counterclockwise* adjacent if C, C' are clockwise adjacent. A chain $C_1, \cdots, C_n$ is called an *adjacent chain* if C_i, C_{i+1} are clockwise adjacent for each $i = 1, \cdots, n-1$ or if, for each i, they are counterclockwise adjacent. We shall also consider infinite collections $\{\cdots, C_{-2}, C_{-1}, C_0, C_1, \cdots\}$ of branched curves with the property that for every $k < m$ the subcollection $C_k, \cdots, C_m$ is a chain. Such a collection will be called an *infinite chain* or *half-infinite chain* depending on whether the indices run through both all positive and all negative integers, or whether there is a least negative or a greatest positive integer. An infinite chain will be called adjacent if every subchain is adjacent.

It is convenient at this point to give some attention to a theorem due to Adkisson and MacLane [I] which states that if $\bar{T}$, $\bar{T}'$ are two homeomorphic Peano continua lying on spheres S, S' respectively, then a homeomorphism from $\bar{T}$ to $\bar{T}'$ can be extended to a homeomorphism of S to S' if and only if it preserves the relative sense of every pair of triods of $\bar{T}$. By a triod, $t = [\alpha, \beta, \gamma]$, of $\bar{T}$ is meant any set of three arcs α, β, γ in $\bar{T}$ which have only a single point, called the vertex, in common. A homeomorphism is said to preserve the relative sense of triods of $\bar{T}$ if every two triods t_1, t_2 which have the same sense (i. e., both clockwise or both counterclockwise) on S are carried into two triods t_1', t_2' of $\bar{T}'$ which have the same sense on S'. Let us denote by $\bar{\pi}$ the plane π plus the point ∞ and by $\bar{T}$ the tree T plus the point ∞. Assuming for the moment that the set $\bar{T}$ is a Peano continuum, the theorem above is applicable to our situation. It is a direct consequence of this theorem that, if T, T' are two homeomorphic trees on π and the xy-plane respectively, then any homeomorphism between them may be extended to a homeomorphism of the planes if and only if the relative sense of the curves of $St(b_1)$, $St(b_2)$ is preserved for every pair of branch points b_1, b_2 of T, the only possible vertices for triods being branch points of T and ∞. In order to show that this is a consequence of the theorem, it must only be shown that the relative sense of every pair of triods of $\bar{T}$ is preserved if this is true for every triod of T, i. e. we must consider triods with vertex at ∞. This will be seen to follow from Theorem 6 of the same paper which states that two non-intersecting triods $t_1 = [\alpha_1, \beta_1, \gamma_1]$, $t_2 = [\alpha_2, \beta_2, \gamma_2]$ have opposite sense on a sphere S if and only if there exists on S a θ-graph whose vertices are the vertices of t_1 and t_2 and whose three (non-intersecting) arcs contain respectively the legs α_1 and α_2, β_1 and β_2, γ_1 and γ_2. Now it is clear from condition (3) in the definition of a tree (the arcwise connectedness) that given any triod with vertex at ∞, it is possible to find at least one triod with vertex at a branch point of T which is, with it, part of a θ-graph, i. e. a triod with sense opposite that of the given one and with vertex at a branch point. Thus if the sense of triods whose vertices are branch points is always preserved, so also for those whose vertex is at ∞. The conclusion is immediate that we may restate the theorem of Adkisson and MacLane, as we have above, for our own purpose here, if it is true that $\bar{T}$ is a Peano space. But this is clearly true: it is connected by condition (3) of the definition; it is locally connected and locally compact since by condition (4) open C and open $St(b)$ are open sets in T, however these sets have these properties and they cover T. Now T is closed in π by definition, and, on $\bar{\pi}$, ∞ is a limit point of T. By adding this point to T to obtain $\bar{T}$ we clearly obtain a closed set, i. e. $\bar{T}$ is in fact the closure of T in $\bar{\pi}$. Hence $\bar{T}$ is a closed, connected subset of the 2-sphere $\bar{\pi}$. It follows that $\bar{T}$ is compact and hence locally compact. Finally $\bar{T}$ is locally connected since a compact continuum cannot fail to be locally connected at just one point [XI] and, of course, the addition of a single point, ∞, to T does not disturb the local connectedness of T. Thus $\bar{T}$ is a Peano space.

In Section **3.4** a numbering system for the curves of a tree is described and it is made clear that given any tree it is possible to construct a homeomorphic rectilinear model of it in the xy-plane, i. e. a model consisting of closed and half-open line segments. Moreover, this may even be accomplished so that any one given chain, either finite or infinite, will have its image in the model on a straight line, e. g. on the x-axis. Finally all this may be done preserving the sense in $St(b)$ for every $b \varepsilon B$ and thus the homeomorphism of our tree onto its model may be extended to a homeomorphism of π onto the xy-plane. Thus any tree, no matter how badly "twisted," may be straightened out, by a homeomorphism of the entire plane, into a rectilinear tree. This result is established in **3.4** entirely independently of the remainder of this section, the proof being deferred for convenience only. In order to avoid long and cumbersome proofs the theorems of this section will be established by free use of this fact plus some slight appeal to intuition as far as rectilinear graphs in the plane are concerned.

We now consider relations between a tree T and its complementary domains. In this connection it is convenient to consider a special class of adjacent chains (of curves of T) which we shall call maximal chains. An adjacent chain of curves of a tree is said to be *maximal* if it is not a subchain of any adjacent chain. It is an immediate consequence of our definitions that a chain of adjacent curves is maximal if and only if (1) it is infinite, or (2) it is half-infinite and its initial (or terminal) curve has only one endpoint, or (3) it is a finite chain and both its initial and terminal curves have each only one endpoint (a curve of a tree with only one endpoint extends to infinity in the direction opposite to that with the endpoint). Moreover, since a maximal chain is closed as a subset of the tree and a tree is a closed subset of π, so also is every maximal chain a closed subset, and thus is in fact an open curve extending to infinity in each direction, thus dividing the plane into two Jordan domains. (Reference to Figure 3 will clarify this section as well as the following one.)

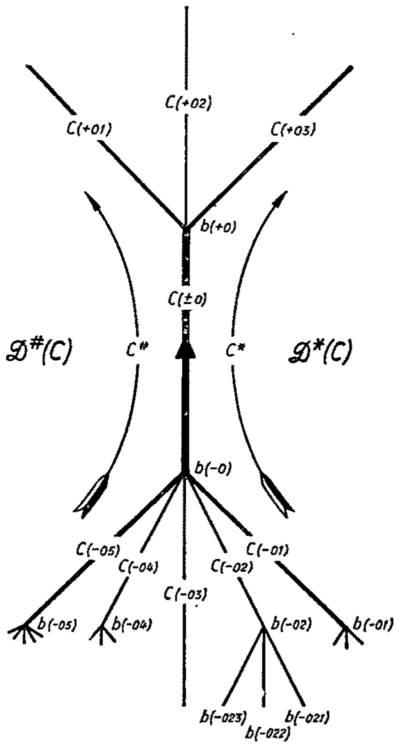


FIGURE 3
RECTILINEAR TREE

THEOREM 3. 3-3. *If T is itself a single curve, then it is its only maximal chain. When T contains more than one curve, then* (1) *each curve of T is contained in exactly two maximal chains which intersect only on this curve and* (2) *every branch point of order n is contained in exactly $2n$ maximal chains whose only common point is the branch point itself. Conversely, the intersection of any two maximal chains can be empty, be a single branch point, or, at most a curve of the tree.*

Proof. Let $C_1, \cdots, C_k$ be any clockwise adjacent chain of two or more curves. Now if C_1 has only one endpoint, then there is no curve C' adjacent to C_1 such that $C', C_1, \cdots, C_k$ is a clockwise adjacent chain, since adjacent curves must have an endpoint in common and at a given endpoint there is exactly one curve clockwise adjacent to a given curve. But, if C_1 has two endpoints, then there is exactly one curve C' such that $C', C_1, \cdots, C_k$ is a clockwise adjacent chain. Similar remarks apply to C_k. If neither C_1 nor C_k has more than a single endpoint, then the chain is maximal; in any other case we may extend the chain, one curve at a time added to the initial or final curve, until we arrive at endcurves which have only one endpoint; or, if we do not come to a curve with one endpoint, indefinitely. In any of these cases, the resulting adjacent chain is maximal since it is an open curve extending to ∞ in both directions. Thus every finite adjacent chain which is not alerady maximal can be extended to a unique maximal chain.

If we begin with a single curve C, with at least one endpoint b, then there is one curve clockwise adjacent to C in $St(b)$ and one counterclockwise adjacent. Thus in $St(b)$ we have C, C' and C, C'', unique adjacent chains containing C, one clockwise and one counterclockwise. Hence, C is contained in just exactly two maximal chains, one of which contains C, C', the other C, C''. Similarly, if b is a branch point, there are just $2n$ pairs of adjacent curves in $St(b)$, whence b is contained in $2n$ maximal chains.

Now consider the converse. If two chains intersect, they surely must have a branch point b in common. If this is their only point of intersection in open $St(b)$, then they can intersect at no other point, since the tree is arcwise connected and can contain no closed polygon. If they intersect along two curves of $St(b)$, they must be adjacent curves since the chains are adjacent chains; hence by the preceding remarks on uniqueness they must coincide. This leaves only the possibility that they intersect along a single curve of $St(b)$, and in this case again, since there are no closed curves in the tree, they either have no other intersection or they coincide.

THEOREM 3. 3-4. *Every maximal chain of a tree T divides the plane* *into two domains, whose complete boundary it is; and one of these domains contains no points of* T.

Proof. The first part of this theorem is just the Jordan curve theorem. The second part becomes clear when we consider a rectilinear model of the tree. Using the Theorem of Adkisson and MacLane and its consequences as mentioned above, we first map π onto the xy-plane so that the maximal chain becomes the x-axis, every curve of T a chain of line segments, and moreover, so that the orientation is preserved, i. e., so that every clockwise adjacent pair of π will still be clockwise adjacent on the xy-plane, and conversely. Now the contention is that all of the image of T, except what is on the x-axis, will lie in one half-plane, say the upper half-plane. If this is not the case, then there will be a point (u, v) of the upper half-plane and a point (x, y) of the lower half-plane, each in the image T' of T. Then, from $C'(u, v)$, the image-curve containing (u, v), there is a chain to any image-curve on the x-axis, i. e., in the given maximal chain. Let C be the last line segment on the last curve of some such chain to lie in the upper half-plane (more precisely, the endpoint p of C lies on the x-axis, but the rest of the curve lies in the upper half-plane). Similarly, we may choose a line segment C' of T' which lies in the lower half-plane except for one endpoint q. For convenience assume p is to the left of q. Clearly, p and q are branch points on that maximal chain of T' which is the image of our maximal chain in π. Now let C_1, C_2, C_3, C_4 be curves of the maximal chain, i. e., line segments on the x-axis, numbered from left to right, such that p is the common endpoint of the first pair, q of the second. Then it is clear that $[C_1, C, C_2]$ is a triod, with vertex p, of curves in clockwise order, and $[C_3, C, C_4]$ is a triod, with vertex q, of curves in counterclockwise order. But this is impossible since C_1, C_2 and C_3, C_4 are both presumed to be adjacent in the same sense, lying as they do on an adjacent chain. Thus all of T' must lie in either the closed upper half-plane or the closed lower half-plane; whence, the theorem is immediate.

Now let C be a *directed* curve of T, a tree consisting of more than one curve. Then we have seen that C determines exactly two maximal curves which we shall denote by C^* and $C^\#$ with the following convention: as we move along C^* in the direction corresponding to the positive direction on C, then the complementary domain of C^* "to the right" (this can clearly be defined in a topologically invariant manner, by the method used in the proof above) will contain no points of T, and as we move along $C^\#$ in the direction corresponding to the positive direction on C, the complementary domain "to the left" will contain no points of T. These domains will be denoted $\mathcal{D}^*(C^*)$ and $\mathcal{D}^\#(C^\#)$ respectively, or more simply by $\mathcal{D}^*(C)$ and $\mathcal{D}^\#(C)$ respectively. Now as proved above, C^* is the common boundary of two Jordan domains, and the notation for one of them was given as $\mathcal{D}^*(C^*)$, the other will be denoted by $\mathcal{D}^\#(C^*)$. Similarly, $C^\#$ divides the plane into the domains $\mathcal{D}^\#(C^\#)$ and $\mathcal{D}^*(C^\#)$. When T is just a single curve then C, C^* and $C^\#$ are all the same curve, and $\mathcal{D}^*(C^\#) = 0 = \mathcal{D}^\#(C^*)$. If we reverse the direction on C, we must replace $\#$ by $*$ throughout.

If we remove *open* C from $C^* \cup C^\#$ we get either two or four half-open arcs extending to infinity from the endpoint(s) of C: two if C has one endpoint, four if it has two. We let $\delta^*(C+)$ denote the arc from the positive endpoint of C lying on C^*, and $\delta^\#(C+)$ the arc from the positive endpoint of C lying on $C^\#$. Similarly, we use the notation $\delta^*(C—)$ and $\delta^\#(C—)$ for the arcs at the other endpoint. We also let $\delta(C+)$ stand for $\delta^*(C+) \cup \delta^\#(C+)$, and $\delta(C—)$ for $\delta^\#(C—) \cup \delta^*(C—)$, and finally, $\delta(C)$ for $\delta(C+) \cup \delta(C—)$. These symbols are clarified by reference to Figure 3.

The collection of all curves C^* and $C^\#$ are then just the maximal chains of T. As already noted above each of these maximal chains bounds two domains, one of which contains no points of T. A converse to this also holds, and, denoting by $\overline{\pi}$ the extended plane and $\bar{T}$ the points of T plus the point at infinity, we have:

THEOREM 3. 3-5. *If T is a tree of π, then $\overline{\pi} — \bar{T}$ consists of an at most countable collection of Jordan domains, each bounded by a simple closed curve in $\bar{T}$ containing the point at infinity. The necessary and sufficient condition that a curve of $\bar{T}$ bound one of those domains is that it be a maximal chain of curves of T.*

COROLLARY 1. *If T is a tree of a regular curve family F, then each complementary domain decomposes into a sum of sets T_C of F.*

COROLLARY 2. *The complementary domains D_n, if infinite in number, tend uniformly to infinity with any sequence p_n, $p_n \varepsilon\, b\, dry\, D_n$, of their boundary points.*

Proof. As we have already noted, π may be mapped on the xy-plane so that the image of T is rectilinear and even so that a given arc (or chain) of T goes onto the x-axis. It is clear then that the complementary domains are Jordan domains. The plane has a countable basis of open sets, whence it follows trivially that the number of complementary domains must be countable also. A boundary curve of a complementary domain must contain the point at infinity, since T contains no closed curves. Finally, it is clear that π may be mapped onto the xy-plane so that a given complementary domain maps onto the upper half-plane and its boundary onto the x-axis. Thus each such boundary must be a maximal chain.

Corollary 1 follows from the fact that each set T_G is connected and disjoint from every other such set, and hence must lie in a single complementary domain. If Corollary 2 were not true we would obtain a contradiction to either property (4) of a tree or the fact that T is a closed subset of π.

3.4. A numbering system for the curves of a tree.

To facilitate further proofs it will be convenient to establish a system for numbering the curves of a tree of π. The numbering proceeds from an arbitrarily chosen, directed curve C of T, which we shall call the *base curve* of the tree. Using the orientation of the plane together with the existence of a unique chain from the base curve C to each curve of T, we set up a 1-1 correspondence between curves of T and a collection of signed finite sequences, the particular collection depending on both T and C, the sign of the sequences depending only on the direction of C. (See Figure 3.)

To the curve C itself we assign, ambiguously, the sequences ± 0, and we write $C = C(\pm 0)$. If C has a positive endpoint, we denote it by $b(+0)$ and number in *clockwise* order, the curves of $St(b(+0))$ by $[C(+0)]$, $C(+01)$, $C(+02), \cdots, C(0u(+0))$, where $u(+0)$ is defined as the number of curves in $St(b(+0))$ less one, i. e., as twice the multiplicity of that branch point less one. We then denote, if it exists, the endpoint of $C(+0k)$ opposite $b(+0)$ by $b(+0k)$. We follow exactly the same procedure at the other endpoint, if there is one, of $C(\pm 0)$. This endpoint is denoted by $b(-0)$ and the curves of $St(b(-0))$ are numbered, again in *clockwise* order, $[C(-0)]$, $C(-01)$, $C(-02), \cdots, C(-0u(-0))$. If the chain $c(C, C')$ from the base curve C to another curve C' of T contains n curves, we shall say that C' is of *order* n with respect to C. The process above then has numbered every curve of T of order 1 or 2 by *exactly one* finite sequence of one or two elements respectively (except for the ambiguity in the numbering of C itself). Moreover, it assigns a unique sequence to the endpoints of the curve, with the endpoint being numbered with the same number as the curve of lowest order having it as endpoint. Two curves C, C' of the same order will be clockwise adjacent if the final integer of the sequence of C is one less than that of the sequence for C'; and two curves C, C' with C of lower order than C' will be clockwise adjacent if the sequence of C' is that of C with a final integer 1 added to it. Finally, the chain from the base curve to a curve C' consists of the curves whose numbering sequences are successive *"lower segments"* of the sequence numbering C', i. e., if $\alpha = 0p_2, \cdots, p_n$ numbers C', then $\alpha_1 = 0$, $\alpha_2 = 0p_2$, $\cdots, \alpha_{n-1} = 0p_2 \cdots p_{n-1}$, $\alpha = \alpha_n = 0p_2 \cdots p_{n-1}p_n$ number the curves of the chain from the base curve to C'.

Now if we assume that everything said above is true for every curve of order n, it is very simple to show that it may be extended *in toto* to the curves of order $n+1$; i. e., let C' be any curve of T of order $n+1$. Then C' is the terminal curve of the chain $c(C, C')$ of $n+1$ curves, all of which except C' itself have already received their unique numbering the next-to-last of them by a sequence α of n terms, which sequence also numbers the common endpoint $b(\alpha)$ of this curve and C'. As before, we number the curves in $St(b(\alpha))$ in clockwise order as $[C(\alpha)]$, $C(\alpha, 1)$, $C(\alpha, 2), \cdots, C(\alpha, u(\alpha))$. Here we let α, k denote the sequence whose first n elements are those of α, and whose $(n+1)$-st is k, i. e., we adjoin k to the sequence α. In this process C' will receive a unique numbering, and the statements above will follow through.

With the help of a little new terminology, we will express these facts in a theorem. Given two collections of sequences, A, A^*, we denote by $A \cup A^*$ the collection of all signed sequences obtained by giving those in A positive sign and those in A^* negative. Using this notation we shall call a collection A of finite sequences *allowable* if:

(1) Every sequence has 0 as first element, positive integers for the other elements, and 0 is a sequence of A.

(2) $\alpha, k \varepsilon A$ implies $\alpha, k-1 \varepsilon A$ if $k \neq 1$, and implies $\alpha \varepsilon A$ if $k = 1$.

(3) For each $\alpha \varepsilon A$ there is defined an integer $u(\alpha) = 0$ or $\geqq 3$ and odd such that if $u(\alpha) \neq 0$ then $\alpha, 1; \alpha, 2; \cdots; \alpha, u(\alpha)$ are in A but not $\alpha, u(\alpha)+1$; and, if $u(\alpha) = 0$, then there is no sequence of A with α as lower segment, i. e., of the form $\alpha, p_{n+1}p_{n+2} \cdots p_{n+k}$. (Note: If $u(\alpha) = 0$ we call α a *terminal* sequence.)

THEOREM 3.4-1. *Given a tree T, a curve C of T, and a direction on C, then there exist two unique allowable collections of finite sequences, A, A^*, such that there is a* 1-1 *correspondence between the curves of T and the signed sequences $A \cup A^*$ (except for ± 0 being assigned to C), and such that there is further a* 1-1 *correspondence between the endpoints of the curves of T and those signed sequences which are not terminal of the collection $A \cup A^*$, these correspondences being as described above and having in particular the properties:*

(1) *If* $C(\alpha)$ *is any curve of* T, *then* $C(\pm 0), C(\alpha_2), \cdots, C(\alpha_{n-1})$, $C(\alpha)$ *is the chain from the base curve to* $C(\alpha)$.

(2) $C(\alpha, k)$ *has the endpoint* $b(\alpha)$ *in common with the lower order curve* $C(\alpha)$ *and, if* α, k *is not a terminal sequence, the endpoint* $b(\alpha, k)$ *at the opposite end.*

(3) $C(\alpha)$, $C(\beta)$ *of the same order* n *are clockwise [counterclockwise] adjacent if and only if* $\alpha_{n-1} = \beta_{n-1}$ *and* $\alpha = \alpha_{n-1}, k$; $\beta = \beta_{n-1}, k+1$ $[\beta = \beta_{n-1}, k-1]$. $C(\alpha)$, $C(\beta)$ *of different order are clockwise [counterclockwise] adjacent if and only if* $\beta = \alpha, 1$ $[\beta = \alpha, u(\alpha)]$.

It is obvious but tedious to prove that maximal chains, the set $\delta^*(C+)$, $\delta^\#(C+)$ and so on are numbered by sequences with certain characteristic properties. We shall not develop this aspect, but shall state one or two important properties of the numbering below:

THEOREM 3.4-2. *Two trees* T *on* π *and* T' *on* π' *are homeomorphic under a homeomorphism which may be extended to all of* π *if and only if we may choose and direct a base curve from each tree and orient the image plane so that the numberings of the two trees are then identical.*

Proof. If the two trees are homeomorphic under such a homeomorphism, then it is clear that for any directed curve C of T as base curve we may choose the homeomorph C' of C as base curve in T', giving it the direction induced by C, and using the orientation in π' induced by the orientation in π, we will get precisely the same numbering for each tree.

On the other hand let T, T' be two trees with identical numberings. We first show that they are homeomorphic. We let $f(\alpha): C(\alpha) \rightarrow C'(\alpha)$ be any homeomorphism of $C(\alpha)$ *onto* $C'(\alpha)$ such that $b(\alpha_{n-1})$ maps onto $b'(\alpha_{n-1})$ or, if $\alpha = 0$, so that $b(0+)$ maps onto $b'(0+)$, then $f(\alpha)$ coincides with $f(\alpha_{n-1})$ at $b(\alpha_{n-1})$, the only point where their domains overlap. The map $f: T \rightarrow T'$ defined by $f(x) = f(\alpha)x$, for α such that $x \in C(\alpha)$, is 1-1, and is continuous on each of a family of closed sets covering T. Now let x_n be any sequence of points on T such that $x_n \rightarrow x \in T$, then by property (4) of trees, for $n \geqq N$, x_n will lie on C_x or $St(x)$, the latter if x is a branch point. From the continuity of f on C_x and $St(x)$ for every $x \in T$, it follows that $f(x_n) \rightarrow f(x)$. Hence, since x_n was any sequence and x any point, f is continuous on T. It follows in the same manner that f^{-1} is continuous on T'. Thus f is a homeomorphism from T to T'.

Now in view of the fact that for every branch point b of T, the sense in $St(b)$ must be preserved by f as defined above, by virtue of our earlier slight modification of the theorem of Adkisson and MacLane, it follows that f can be extended to a homeomorphism on all of π.

We now remark that this theorem makes it possible to construct a rectilinear model of any tree T on the xy-plane, and assures us that there will be a homeomorphism of π onto the xy-plane carrying T onto this model. The model is constructed by considering any numbering $A \cup A^*$ of T, and, using line segments of length $\geqq 1$ as our elements, building up the model piece by piece: We begin with a base element corresponding to the sequence ± 0, add segments corresponding to the 2nd order sequences, 3rd order, etc., each time moving further out from our base segment so that its distance from any n-th order segment approaches infinity with n. In this process it is clearly possible to construct the model so that the image of any one particular chain is a straight line, e. g., the x-axis. (See Figure 3.)

3. 5. r-sets and cross-sections. For an arc pq, lying on an adjacent chain of curves $C_1, \cdots, C_n$, it is possible to get a serviceable analog of the r-neighborhood of an arc on a regular curve (cf. Theorem 1. 2-2). By suitably directing C_1, we have both pq and $C_1, \cdots, C_n$ as arcs on $C_1{}^*$. By an *r-set of pq* will be meant any closed set $U \subset \mathfrak{D}^*(C_1{}^*)$ together with a homeomorphism g of U onto the rectangle R_1 of the xy-plane, where $R_1 = \{(x, y) \mid -1 \leqq x \leqq 1, 0 \leqq y \leqq 1\}$, with g having the properties:

(1) g carries $F[U]$ onto the lines $y = \text{constant}$ in R_1.

(2) $g^{-1}(\lambda_i)$ are cross-sections, where λ_i, $i = 1, 2$ are, respectively, that part of the lines $x = -1$ and $x = +1$ in R_1.

(3) pq is mapped into the set $\{(x, y) \mid y = 0, -1 < x < 1\}$.

$U(pq)$ contains no branch points except those on $C_1{}^*$ itself and lies entirely in $\mathfrak{D}^*(C_1{}^*) \cup C_1{}^*$, which contains no points of T_{C_1} except those on $C_1{}^*$. (See Figure 4.) In the event that pq lies on a single curve C_1, then two r-sets which coincide along C_1 but lie on different sides of C_1 will together form an r-neighborhood of C_1. We shall find it convenient to refer to an r-set of a single point p, by which we shall mean one side of a regular neighborhood of p if p is a regular point and one sector of an admissible neighborhood if p is a branch point. We now prove an analog of part of Theorem 1. 2-2.

THEOREM 3. 5-1. *Given any arc pq on an adjacent chain of curves $C_1, \cdots, C_n$, $n > 1$ and any $\epsilon > 0$, there exists an r-set of pq which lies inside the ϵ-neighborhood of pq and in that complementary domain of T whose boundary is the maximal chain containing the given adjacent chain.*

Proof. Let C_1 be directed so that C_1^* contains the chain $C_1, \cdots, C_n$. We leave F unchanged in $\mathfrak{D}^*(C_1^*)$, the complementary domain of T_{C_1} in which U is to lie, but we alter F in the remainder of the plane by mapping the lines $y = \text{constant}$ of the lower half-plane, including the x-axis onto $C_1^* \cup \mathfrak{D}^\#(C_1^*)$ in such a manner that the x-axis is mapped onto C_1^*. Then by Theorem 1. 2-3 this new family is regular at all points of C_1^* and agrees with F in $C_1^* \cup \mathfrak{D}^*(C_1^*)$. Hence by Theorem 1. 2-2, there is an arbitrarily small r-neighborhood of pq, call it V. Then $U(pq) = \overline{V \cap [C_1^* \cup \mathfrak{D}^*(C_1^*)]}$ will be our desired r-set.

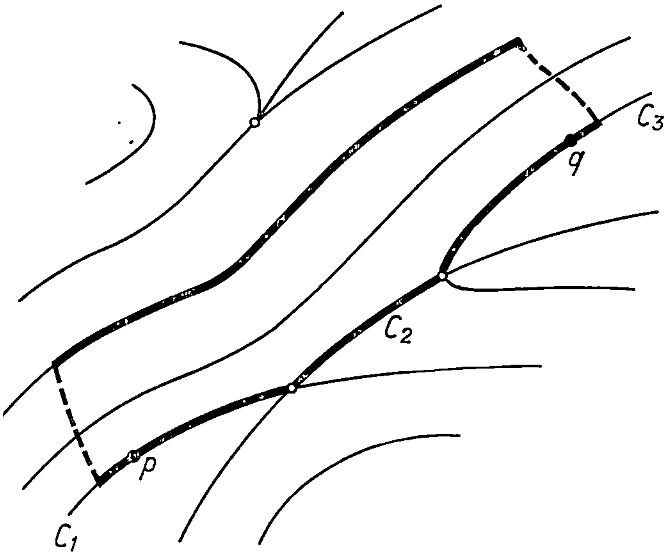


FIGURE 4 AN R-SET

THEOREM 3. 5-2. *Let a sequence of points q_n on distinct curves C_n approach the point q, where q is a regular point or a branch point, then there is a curve C which may be so directed that q lies on C^* and an infinite subsequence q_m of q_n lies in $\mathfrak{D}^*(C^*)$. Moreover, if p is any point on C^*, then there is a sequence of points p_m in $\mathfrak{D}^*(C^*)$ such that $p_m \to p'$, and such that for $m > M$ the curves C_{p_m} and C_{q_m} are identical.*

Proof. Let T be the tree of F which contains q, then there are at most a finite number of complementary domains of T on whose boundary q lies, and since the q_n lie on distinct curves, there must be a subsequence q_m of these points lying in one of these complementary domains. The maximal chain which bounds this domain, can, for some suitably directed C be given as $\mathfrak{D}^*(C^*)$. Now let p be any other point on C^*. We may take an r-set $U(pq)$ of pq in $\mathfrak{D}^*(C^*)$ and, if f denotes the homeomorphism $f: U(pq) \rightarrow R_1$, then the curves C_m containing q_m map onto lines $y = k_m$ for $m \geqq M$. If $f(p) = (x, 0)$, then the points $p_m = f^{-1}(x, k_m)$ will be the desired sequence.

THEOREM 3. 5-3. *An arc pqr is a cross-section of F if and only if* (1) *it contains no branch points,* (2) *one of the domains* $\mathfrak{D}^*(C_q)$, $\mathfrak{D}^\#(C_q)$ *contains p, the other r, and* (3) *pq and qr are each cross-sections.*

Proof. We first assume that the arc pqr is a cross-section through q. Then (1) and (3) follow by definition of cross-section. By the Lemma stated in [IV, p. 158] there is an r-neighborhood of q, $V(q)$, such that the image of pqr in R_0 is the y-axis. Every curve crossing $V(q)$ crosses pqr; hence, no curve has more than one line $y =$ constant as image in R_0. The point q itself maps on $(0, 0)$ and C_q on the x-axis; hence, $V - C_q$ splits into two domains, one containing p, whose image in R_0 is $(0, 1)$, and the other r, whose image in R_0 is $(0, -1)$. Moreover, one of these domains lies in $\mathfrak{D}^*(C_q)$ and the other in $\mathfrak{D}^\#(C_q)$, for q is a point on the common boundary of these two domains and hence every neighborhood of q contains points of each domain.

Now, if we assume that pqr is an arc with the properties (1), (2) and (3), we may show that it is a cross-section by showing that it intersects any set T_C at most once. This is clear at once if we remember that a set T_C cannot have points in each of the domains $\mathfrak{D}^*(C_q)$ and $\mathfrak{D}^\#(C_q)$ so that if pqr had more than one point in common with T_C, each such common point would have to lie in the same domain, i. e., both on pq, or both on qr. This is impossible, however, since both of these arcs are cross-sections. It follows that pqr is a cross-section.

The following corollary is immediate:

COROLLARY 1. *If C, C' both intersect a cross-section pq, and each is directed to cross pq in the same direction, then either* $\mathfrak{D}^*(C) \supset \mathfrak{D}^*(C')$ *or* $\mathfrak{D}^\#(C) \subset \mathfrak{D}^\#(C')$.

COROLLARY 2. *If an arc is such that any point on it is interior to a subarc which is a cross-section, then it is a cross-section.*

Proof. Let pq be such an arc; then we may cover pq with a finite number of r-neighborhoods which overlap. Then, applying the theorem repeatedly a finite number of times gives the result desired.

4. The Family F as the Level Curves of a Continuous Function.

In this section it will be shown that there is a continuous function $f(x, y)$ whose level curves are exactly the family F. The proof of this statement will depend on our ability to remove certain branched curves of F together with their endpoints so as to leave a subset R^* of the plane π, which is open, connected, and simply connected and is such that $F^* = F[R^*]$ is a regular curve family filling R^*. It will then follow from [IV] that there is a continuous function $f^*(x, y)$ defined on R^* and having the family F^* as level curves. Finally, it is shown that $f^*(x, y)$ may be extended to a continuous function on all of the plane with the curves of F as level curves. In this and the next section we will restrict the use of the term *tree* to those sets T_C containing singular curves, i. e. the term will not be used to include a regular curve.

4. 1. The numbering of the trees of F.

THEOREM 4. 1-1. *If K is any compact subset of π, then there are at most a finite number of distinct trees of F which intersect K on more than one curve of the tree. Moreover, no more than a finite number of curves from any one tree can intersect K.*

Proof. The second part of the theorem is an immediate consequence of the fact that any point p which is a limit of a sequence of points p_n of a tree must be a point of the tree and, in addition to this, for $n \geqq N$, must lie in $St(p)$ if p is a branch point, or on C_p if p is a regular point. If an infinite sequence of curves of a single tree intersected K, we could, by compactness of K, choose a sequence of points on *distinct* curves of this tree such that the sequence would have a limit point, and hence could not conform to the requirements above.

We prove the first part of the theorem by assuming it false and arriving at a contradiction. Let T_i $(i = 1, 2, \cdots)$ be an infinite collection of trees, each intersecting the compact set K on at least the two curves C_i, C_i'. Using the compactness of K we may choose an infinite subsequence from this collection of trees together with points $p_i \varepsilon C_i \cap K$ and $q_i \varepsilon C_i' \cap K$ on each tree T_i of the subsequence such that $p_i \rightarrow p$ and $q_i \rightarrow q$. We then throw away all trees except those of our subsequence and renumber. We may assume p is a regular point lying on a curve C and that the p_i all lie in a single complementary domain of T_C, i. e. in $\mathfrak{D}^*(C)$. If this is not the case, using Theorem 3. 5-2 we may rechoose p and p_i so that they satisfy these conditions.

This may move them out of K, but this causes no difficulty later. We still have $p_i \varepsilon C_i \subset T_i$. In this process, to get the p_i all in a single complementary domain of T_C, it may be necessary to choose a subsequence. If so we throw out the trees T_i containing the points p_i which we remove to get our subsequence, then we renumber. That is, we simultaneously pass to subsequences of p_i and of q_i which correspond, so that we still have $p_i \varepsilon C_i$, $q_i \varepsilon C_i'$ and and C_i, $C_i' \subset T_i$ after renumbering. This method is followed whenever it is necessary to select a subsequence of either p_i or q_i and renumber. Now we may similarly assume q is a regular point on a curve C' and that the q_i lie in $\mathfrak{D}^*(C')$, or else use Theorem 3.5-2 as above, and again choose a subsequence. When this process is finished we may be sure that $p \neq q$, for if this were not the case, choosing an r-neighborhood of $p = q$, we would find it crossed by curves C_i, C_i' of the same tree T_i, and this is impossible since it would imply that a cross-section crossed T_i twice.

Now a brief consideration of the image in R_0 of non-intersecting r-neighborhoods $U(p)$ and $U(q)$ and the images of p_i and q_i which lie in these neighborhoods for $i > N$, makes it obvious that we may choose three trees from our sequence of trees: T_{i_1}, T_{i_2}, T_{i_3} which are such that p_{i_1}, p_{i_2}, p_{i_3} lie in $U(p)$ and may be connected in that order by a cross-section γ running from p_{i_1} through p_{i_2} to p_{i_3}; and at the same time q_{i_1}, q_{i_2}, q_{i_3} lie in $U(q)$ and similarly may be connected by a cross-section γ' running from q_{i_1} through q_{i_2} to q_{i_3} (reference to Figure 5 will make the procedure here more easily followed). Then, denoting by c_j the chain $c(C_{i_j}, C_{i_j}')$ in T_{i_j}, $j = 1, 2, 3$, we may define the following three arcs, λ_{i_1}, λ_{i_2}, λ_{i_3} having only their endpoints p_{i_2} and q_{i_2} in common: λ_{i_2} is the arc $p_{i_2}p_{i_1}$ on γ plus the arc $p_{i_1}q_{i_1}$ on c_1 plus the arc $p_{i_1}q_{i_2}$ on γ'. λ_{i_3} is similarly defined with 1 replaced by 3 in the subscripts above, and finally λ_{i_2} is the arc $p_{i_2}q_{i_2}$ on c_2. Two of these arcs, say λ_{j_1}, λ_{j_2}, must form a simple closed curve Γ containing the third λ_{j_3} in its interior. But this is impossible, since each arc contains a branch point; in particular the arc λ_{j_3}, thus enclosed in the interior (in our example) would contain a branch point; and from this branch point issues a chain of curves of T_{j_3}, all distinct from λ_{j_3}, which must leave Γ at some point r. This point r cannot be on T_{j_1} or T_{j_2} since two trees cannot intersect, nor can it be on γ or γ' since then this cross-section would have two points on the same tree, which is ruled out by Theorem 3.3-1. Hence, we conclude that our initial assumption is impossible and that the theorem must be true.

This theorem will be used to give a method of numbering all non-trivial trees, i. e., trees containing singular curves. We choose any regular point p on π and let K_n designate the circle (with its interior) of center p and radius n. Now the number of trees cutting K_1 is, of course, denumerable and we number them in any order as $T_{11}, T_{12}, T_{13}, \cdots$ and choose from each a curve C_{11}, C_{12}, C_{13}, respectively, which itself intersects K_1. By the above theorem these choices of curves will be unique for all except a finite number of the trees, and for these C_{1j} is chosen at random from any one of the finite number of curves of the tree cutting K_1. Next we number the trees which intersect K_2 but *not* K_1 as $T_{21}, T_{22}, T_{23}, \cdots$ etc., and let $C_{21}, C_{22}, C_{23}, \cdots$ respectively be curves of these trees which themselves intersect K_2. Proceeding with this process we number the trees cutting K_n but not K_{n-1} as $T_{n1}, T_{n2}, T_{n3}, \cdots$ etc., and choose from each curves $C_{n1}, C_{n2}, \cdots$ cutting K_n. This process will clearly number all the trees of F, and we choose the curves C_{ij} as base curves of the trees, hence determining within each tree T_{ij} a numbering of its curves by sets of finite sequences $A_{ij} \cup A_{ij}^*$ as described in **3.4.** Our method of numbering the trees guarantees that for $m > n$ no tree T_{mj} intersects K_n and, moreover, for each n, there are at most a finite number of curves of the set $\bigcup_{m,j;\, m \leqq n} (T_{mj} - C_{mj})$ which intersect K_n. For future reference we shall call the above method of numbering trees a *standard* numbering of the trees of F.

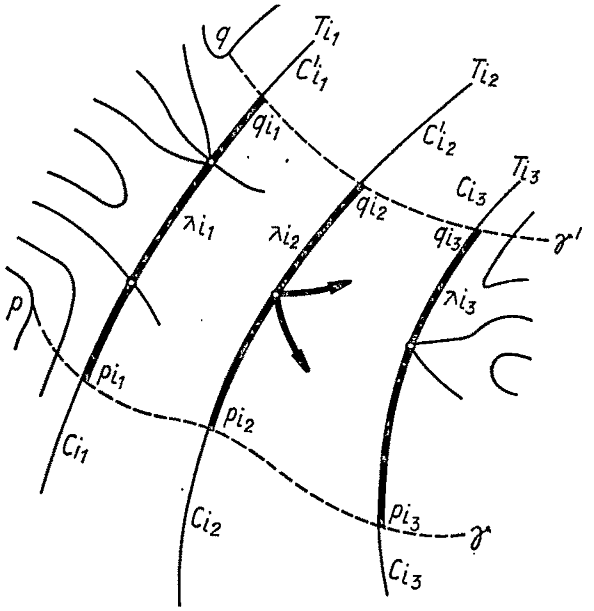


FIGURE 5

With these preliminaries we are able to define the curves which we are going to remove from each tree in order to make the multiply connected region $R = \pi - B$ simply connected. Let T be any tree, then it is numbered say as T_{ij} in a standard numbering as described above, and the choice of base curve C_{ij} included in the standard numbering determines a numbering of the curves of $T = T_{ij}$. Let $b(\alpha)$ be any branch point of this tree with sequence $\alpha = e0k_2 \cdots k_{n-1}k_n$ with $k_n \neq 1$ (where e denotes the sign $+$ or $-$). Any such $b(\alpha)$ will be called the *initial point* of a cut, and the *cut*, $\lambda(b)$, will consist of all curves $C(\alpha, 1), C(\alpha, 1, 1), C(\alpha, 1, 1, 1), \cdots$ and so on *ad infinitum*, or until a terminal sequence $\alpha, 1, 1, \cdots, 1$ is reached, i. e., each cut is a chain of adjacent curves extending from $b(\alpha)$ to infinity. We assume endpoints of the curves included, of course, as part of the cut; thus each cut is of the form $\delta^*[C(\alpha)+]$ or $\delta^\#[C(\alpha)-]$, the former if $b(\alpha)$ is the positive endpoint of $C(\alpha)$ and the latter if it is the negative (see heavy lines in Figure 3). Each $\lambda(b)$ is, again, an arc from $b(\alpha)$ to infinity and includes all branch points numbered by sequences of the form $\alpha, 1, 1, \cdots, 1$. It is clear from the definition of the cuts and the properties of the numbering that every branch point of the tree is on one and only one cut $\lambda(b)$ and that no two cuts intersect at any point. We denote the collection of all the half-open arcs $\lambda(b)$ on T by $\tilde{T}$, and by $\tilde{\mathfrak{T}}$ the sum of the sets $\tilde{T}$ over all the trees of F. The set $R^* = R - \tilde{\mathfrak{T}}$ contains no branch points and is a union of curves of F. Let F^* denote the family $F[R^*]$ filling R^*.

THEOREM 4. 1-2. *R^* is an open, arcwise connected, and simply connected domain, and F^* is regular in R^*.*

Proof. Let q be any point of R^* and let K_n be the first circle with center at p (in the standard numbering scheme) which contains q in its interior. Now consider how much of K_n is removed when $\tilde{\mathfrak{T}}$ is subtracted from π. None of the base curves C_{ij} is in $\tilde{\mathfrak{T}}$ since none of them is in a set $\lambda(b)$ for these curves are assigned the sequence ± 0 in the numbering, which sequence is not of the form $\alpha, 1, \cdots, 1$. And by Theorem 4. 1-1 there can then be at most a finite number of other curves (than base curves) of any T_{ij} in K_n. Hence there is surely an r-neighborhood of q in $K_n - \tilde{\mathfrak{T}}$ and R^* is therefore open and F^* regular.

We wish to show that R^* is arcwise connected. Since every point has an r-neighborhood in R^*, it is clear that R^* is locally-connected. Hence, if it is connected, it is arcwise connected. Now since each cut $\lambda(b)$ is an arc, extending from a point b to infinity, the set $\tilde{\mathfrak{T}} \cup \infty$ on the extended plane $\bar{\pi}$ can clearly be deformed continuously along itself to a single point, the point at infinity. It follows from Eilenberg [III, Theorem 6, p. 77], that R^* is connected.

Finally, if K is any closed curve in R^* containing a point q of $\tilde{\mathfrak{J}}$, then q lies on a cut $\lambda(b)$ which extends to infinity from b and hence must intersect K, contrary to the assumption that K is in R^*. Thus R^* is simply connected.

THEOREM 4.1-3. *Let $B' \subset B$ be the set of all initial points of cuts $\lambda(b)$, then we may define a collection of disjoint, open sets $\{V_b\}$, one for each $b \varepsilon B'$, each containing the corresponding $\lambda(b)$.*

Proof. Referring to the closed circular discs K_n of our standard numbering of the trees of F, we have noted already that only a finite number of the cuts $\lambda(b)$ intersect any K_n. We denote by $B_n' = \{b_j^n\}$ the finite subset of B' whose elements b_j^n are for $j = 1, \cdots, j_n$ those initial points of cuts which intersect K_n but not K_{n-1}. Now, using the normality of π we are able to find disjoint open sets covering the disjoint closed sets $\lambda_1(b_j^1) = \lambda(b_j^1) \cap K_1$. We define $V_1(b_j^1)$ as the intersection of the so chosen open sets covering $\lambda_1(b_j^1)$ with the *interior* of K_1. Then we find disjoint open sets covering each of the closed sets $\lambda_2(b_j^i) = \lambda(b_j^i) \cap [K_2 - \operatorname{int} K_1]$, $i = 1, 2$ and such, moreover, that the open sets covering $\lambda_2(b_j^2)$ do not intersect K_1. Finally, we define $V_2(b_j^i)$, $i = 1, 2$ as the intersections of these open sets with the interior of K_2. Then the sets $V_2(b_j^2)$ and $V_1(b_k^1) \cup V_2(b_k^1)$ are non-intersecting open sets lying in the interior of K_2 and covering $\lambda(b_j^i) \cap int(K_2)$, for all b_j^i's in B_1' or B_2'. This process is continued indefinitely, covering, for each $\lambda(b)$, its intersection with every K_n and in such a manner that none of the open sets covering one cut has a common point with those covering another cut. Then, given any $b \varepsilon B$ it will be in B_n' for some n, hence will be of the form b_i^n, and the cut $\lambda(b_i^n)$ with it as initial point is covered by $V(b_i^n) = \bigcup_{j=n}^{\infty} V_j(b_i^n)$.

THEOREM 4.1-4. *Let F be a branched regular curve family filling the plane π. Then there exists a function $f(p)$ such that:*

(1) *$f(p)$ is defined and continuous for all p in π.*

(2) *for every real number k the locus $f(p) = k$, if not empty, consists of a finite or an at most countably infinite collection of trees (including regular curves) of F.*

(3) *in every neighborhood of any point p in π there are points q for which $f(q) > f(p)$ and points r for which $f(r) < f(p)$.*

Proof. We assume a standard numbering of the non-trivial trees of F and that thus the cuts $\lambda(b)$ and the sets $\tilde{\mathfrak{J}}$ and R^*, etc., are determined. Now this theorem was proved in [IV] by W. Kaplan for curve families regular throughout an open, simply connected domain; thus we may assume that there is a function $f^*(p)$ defined and continuous in R^* and with the properties above. We must show that this function can be extended to a function $f(p)$ with properties (1)-(3) above. The proof, which follows, is divided into three sections, A, B, and C.

(A) First it is necessary to prove that, given any tree T of F, the value of f^* is the same on each curve of $T[R^*]$, i. e., on all curves of T which lie in R^*. Let $C(\pm 0)$ be the base curve of T in the numbering; we shall proceed by induction on the order of the curves of T. If $C(\pm 0)$ has no endpoint, then it is a regular curve, lies entirely in R^* and the result is trivial. Assume it has a positive endpoint $b(+0)$. Then $C(+01)$ is in $\lambda(b(+0))$ and hence not in R^* or $T[R^*]$, but the other curves of $St(b(+0))$ are all in $T[R^*]$. To prove that $f^*(p)$ has the same value on each of these it is only necessary to prove that it has the same value on each pair of adjacent curves among them, for then the value of f^* on $C(+02)$ is the same as that on $C(+03)$ and so on until finally we have the value of $C(+0u(+0))$ the same as that on $C(+0)$. It is quite obvious that this must be so, however, for if C, C' are adjacent curves of $T[R^*]$ and $p \varepsilon C$, $q \varepsilon C'$, then there is an r-set $U(pq)$ in R^*; and, if $p_n \varepsilon U$ is a sequence of points approaching p, then there is a sequence $q_n \varepsilon C_{p_n}$ with $q_n \varepsilon U$ and q_n approaching q. But, since $q_n \varepsilon C_{p_n}$ we have $f^*(q_n) = f^*(p_n)$ and hence $f^*(q) = \lim_{n\to\infty} f^*(q_n) = \lim_{n\to\infty} f^*(p_n) = f^*(p)$. This same procedure actually tells us even more, i. e., that if $C_1, \cdots, C_n$ is any chain of adjacent curves with both $C_1, C_n \subset R^*$, then f^* must have the same value on C_1, C_n.

Now let $C(\alpha, k)$ be a curve of $T[R^*]$ whose sequence is positive and of order $n+1$, and assume that f^* has the same value on each curve of $T[R^*]$ numbered by a positive sequence of order n or less. The sequence α is of the form $\alpha = 0k_2 \cdots k_{n-1}k_n$; and we consider two cases: (1) $k_n \neq 1$ and (2) $k_n = 1$; in either event $k \neq 1$ since $C(\alpha, k)$ is in R^*. In case (1) $b(\alpha)$ is the initial point of a cut, hence $C(\alpha, 1)$ is the only curve of $St(b(\alpha))$ in the cut, and thus the only curve of $St(b(\alpha))$ not in R^*. Moreover, $St(b(\alpha))$ contains the curve $C(\alpha)$ of order n. It follows by precisely the same argument as above that f^* has the same value on each of the curves of $St(b(\alpha))$ in R^* and, in particular on $C(\alpha, k)$, as it has on the n-th order curve $C(\alpha)$ and hence that it has on $C(\pm 0)$. In case (2) both the curves $C(\alpha)$ and

$C(\alpha, 1)$ of $St(b(\alpha))$ are in a cut, i. e. not in R^*. But the curves $C(\alpha_{n-1}, 2)$, $C(\alpha_{n-1}, 1) = C(\alpha)$, and $C(\alpha, u(\alpha))$ form an adjacent chain with the first and last curves in R^*. On the first curve f^* has the same value as on $C(\pm 0)$ since it is of order n, hence it has this value also on the last, $C(\alpha, w(\alpha))$. Now, by going from adjacent curve to adjacent curve, we see that this must be the value of f^* on each curve of $St(b(\alpha))$ in R^* and in particular on $C(\alpha, k)$. This completes the first step in the proof.

(B) Next we define $f(p)$ at every point of π as follows: $f(p) = f^*(p)$ for $p \varepsilon R^*$, and $f(p) =$ value of f^* on $T[R^*]$ for $p \varepsilon T$. The single-valued function $f(p)$ will then be continuous at each point of R^* (since R^* is an open subset of π and thus the extension cannot affect the continuity of f in that domain), and will have the same value at every point of any tree. Now every point of $\mathfrak{J} = \pi - R^*$ lies on a cut $\lambda(b)$, which in turn lies in a neighborhood $V(\lambda(b))$ not containing points of any other cut. What must be shown is that $f(p)$ is continuous at an arbitrary point q of an arbitrary cut $\lambda(b)$. Now let q_n be any sequence of points approaching the point q of $\lambda(b)$; it must be shown that $f(q_n) \rightarrow f(q)$. We shall denote by T the tree containing q; then since $f(p)$ is constant on T, we shall assume that each q_n lies on a distinct curve and none of them is in T. This involves no loss of generality since the result is trivial otherwise. Moreover, we may restrict ourselves to sequences lying in a single complementary domain of T, the reason being that since the number of complementary domains of T with q as a boundary point is finite, any such sequence q_n, with $q_n \not\varepsilon T$, can be decomposed into a *finite* number of distinct subsequences, containing all the terms of q_n, but no two having a term in common, and each consisting of points from only one complementary domain; and, finally, if for each of these subsequences we have $f(q_{n_i}) \rightarrow f(q)$, then $f(q_n) \rightarrow f(q)$. Thus we need now to consider only a sequence $q_n \rightarrow q$ such that for some $C^* \supset q$, $q_n \varepsilon \mathfrak{D}^*(C^*)$ for all n. C^* is then in T, and since no cut separates π, there is a curve C' on C^* which is in R^*. Let p be any point of C' and $U(qp)$ an r-set of qp in $\mathfrak{D}^*(C^*)$. Then, by Theorem 3. 5-2 there is in U a sequence $p_n \rightarrow p$ with $C_{p_n} \equiv C_{q_n}$ and hence $f(p_n) = f(q_n)$. But p is in R^* and $f(p)$ is continuous in R^*, therefore $\lim f(q_n) = \lim f(p_n) = f(p)$. But this is exactly what is needed, for p, q are both on T and hence $f(p) = f(q)$, so f is continuous at q. This establishes property (1) of the theorem.

Property (2) of the theorem is trivial for $f(p)$, since it is satisfied by f^* in R^* and we have added only a denumerable number of curves to the domain of f^* to get the domain of f.

(C) Finally we must prove property (3), i. e., that $f(p)$ has no weak relative extrema. This is clearly equivalent to the following, at least for regular points: if p is a regular point, then f takes a different value on every curve of every r-neighborhod of p, or again equivalently, is monotone on every cross-section. Then, since any arc pq on a curve C has an r-neighborhod, a function satisfies property (3) at every point of a curve or no point of a curve. As to branch points, we can show at once that the condition is satisfied there, for there is always a curve of $St(b)$ in R^*, hence in any neighborhood of b we may find a point q of a curve of R^* and a neighborhood of this point q inside that of b. Now $f(q) = f(b)$ and in this neighborhood of q there will be points q_1, q_2 such that $f(q_1) < f(q) < f(q_2)$, since we are in R^*, where we know f to have property (3). Since q_1, q_2 are in the given neighborhood of b, we have proved our contention for the case of branch points.

Now we wish to show that if f has property (3) on every curve of $St(b)$ except one, C, where b is any branch point, then f has property (3) on C also. Let the curves of $St(b)$ be numbered counterclockwise $C = C_1, C_2, \cdots, C_{2m}$, m being the order of the branch point b. In $U(b)$, an admissible neighborhood, we shall let s_i denote any arc into the sector bounded by C_i, C_{i+1}, such that s_i without b, its endpoint, is a cross-section, e. g., in the image of U on $|z| < 1$ we could take for s_i radii into the respective sectors. Then we indicate by s_i^+ that f increases as we move from b on s_i, by s_i^- that f decreases. Clearly C_i has property (3) if and only if s_{i-1}^+ implies s_i^- and s_{i-1}^- implies s_i^+. Hence if we have s_1^-, then we have by induction s_j^+ for even j, and in particular s_{2m}^+, whence C_1 has property (3); similarly for s_1^+.

Now let the curves of any cut $\lambda(b)$ be numbered $C_1, C_2, \cdots$ beginning with the initial curve and proceeding out from b. C_1 is the only curve of $St(b)$ not in R^*, and hence it must have property (3). If the n-th curve C_n has property (3) then C_{n+1} is the only curve of $St(C_n \cap C_{n+1})$ not having this property since the other curves (than C_n) are in R^*, thus C_{n+1} also must have the desired property. This proves by induction that every curve of every cut has property (3) and hence so has $f(p)$ for all points of π.

A regular curve family F will be said to be *orientable* if its curves may be directed so that every regular point has an r-neighborhood in which the image of each curve in R_0 is directed toward increasing x. Then we have at once:

COROLLARY. *The branched regular curve family F is orientable as a regular curve family in $R = \pi - B$.*

Proof. Exactly as in W. Kaplan [IV, Remark 2, p. 184-5].

NORTHWESTERN UNIVERSITY.

BIBLIOGRAPHY.

[I] V. W. Adkisson and S. MacLane, "Extendnig maps of plane Peano Continua," *Duke Mathematical Journal*, vol. 6 (1940), pp. 216-228.

[II] Y. S. Chin, *Regular families of curves and pseudo-harmonic functions*, Thesis Harvard University, 1947.

[III] S. Eilenberg, "Transformations continues en circonférence et la topologie du plan," *Fundamenta Mathematicae*, vol. 26 (1936), pp. 61-112.

[IV] W. Kaplan, "Regular curve-families filling the plane, I," *Duke Mathematical Journal*, vol. 7 (1940), pp. 154-185.

[V] ———, "Regular curve-families filling the plane, II," *ibid.*, vol. 8 (1941), pp. 11-46.

[VI] ———, "The structure of a curve-family on a surface in the neighborhood of an isolated singularity," *American Journal of Mathematics*, vol. 64 (1942), pp. 1-35.

[VII] ———, "Differentiability of regular curve-families on the sphere," *Lectures in Topology*, Ann Arbor (1941), pp. 299-301.

[VIII] ———, "Topology of level curves of harmonic functions," *Transactions of the American Mathematical Society*, vol. 63, No. 3 (1948), pp. 514-522.

[IX] B. v. Kerékjártó, *Vorlesungen über Topologie*, I, Berlin (1933).

[X] H. Kneser, "Reguläre Kurvenscharen auf den Ringflächen," *Mathematische Annalen*, vol. 91 (1924), pp. 135-154.

[XI] G. T. Whyburn, *Analytic Topology*, American Mathematical Society Colloquium Publications, vol 28, New York (1942).

ON TOTAL NILPOTENT ALGEBRAS.*

By Roy Dubisch and Sam Perlis.

Introduction. A *total nilpotent algebra* of degree n over $\mathfrak{F}$, denoted by $\mathfrak{F}^{(n)}$, is defined to be any isomorphic copy of the algebra of $n \times n$ matrices over $\mathfrak{F}$ with zeros on and above the diagonal. Interest in this class of algebras stems from the following fact (proved in Section **1**).

THEOREM 1. *The total nilpotent algebras together with their subalgebras constitute the totality of* (*associative*) *nilpotent algebras.*

The quantities of $\mathfrak{F}^{(n)}$ may be regarded as linear transformations on the n-dimensional vector space V over $\mathfrak{F}$. Then (Theorem 2) $\mathfrak{F}^{(n)}$ is characterized isomorphically as a maximal algebra of nilpotent linear transformations on $V/\mathfrak{F}$.

The main theme of our investigation of $\mathfrak{F}^{(n)}$ is the explicit determination of all its automorphisms and ideals. The automorphism group $\mathfrak{G}$ is found to have the form $\mathfrak{G} = \mathfrak{D}\mathfrak{M}$, where $\mathfrak{M}$ is a normal subgroup which is "almost all" of $\mathfrak{G}$, and $\mathfrak{D} \cap \mathfrak{M} = I$. Also, $\mathfrak{M} = \mathfrak{J} \times \mathfrak{N}$ where $\mathfrak{J}$ is the group of inner automorphisms (Section 2) and the group $\mathfrak{N}$ consists of all automorphisms carrying every absolute right or left divisor of zero into itself.

After the ideals are determined (Section 7), attention is turned to the characteristic ideals—those which are carried into themselves by every automorphism of $\mathfrak{F}^{(n)}$—and all of these are determined. Certain of the characteristic ideals form a class of algebras which, taken abstractly, are called elementary nilpotent algebras. All total nilpotent algebras are elementary. The final section begins the study of elementary nilpotent algebras.

1. Maximality properties.

A total nilpotent algebra $\mathfrak{F}^{(n)}$ has a basis of matric units e_{ij} with $1 \leqq j < i \leqq n$ and is generated by those e_{ij} which lie in $\mathfrak{A} = \mathfrak{F}^{(n)}$ but not[1] in $\mathfrak{A}^2$, namely the quantities

$$e_{21}, e_{32}, \cdots, e_{n,n-1}. \tag{1}$$

Thus, $e_{ij} = e_{i,i-1}e_{i-1,i-2} \cdots e_{j+1,j}$. The product of the $n-1$ quantities (1)

---

* Received February 14, 1950; presented to the American Mathematical Society, August 31, 1949.

[1] The reader can verify that e_{ij} lies in $\mathfrak{A}^k$ if and only if $i - j \geqq k$. Hence e_{ij} lies in $\mathfrak{A}^k$ but not in $\mathfrak{A}^{k+1}$ if and only if $i - j = k$.

in reverse order is $e_{n1} \neq 0$, but every product of n factors chosen from the list (1) is zero. It follows that $\mathfrak{A}^{n-1} > \mathfrak{A}^{n} = 0$, that is, $\mathfrak{F}^{(n)}$ is nilpotent of index n. To complete the proof of Theorem 1, let $\mathfrak{B}$ be any nilpotent algebra of order $n-1$ and let $\mathfrak{B}_1$ denote the algebra $(1, \mathfrak{B})$ obtained by adjunction of a unity to $\mathfrak{B}$. The regular representation of $\mathfrak{B}_1$ contains a subalgebra $\mathfrak{B}_0$ isomorphic to $\mathfrak{B}$. If in obtaining the regular representation we use a normalized basis[2] of $\mathfrak{B}$, we find[3] that the matrices in $\mathfrak{B}_0$ belong to $\mathfrak{F}^{(n)}$.

THEOREM 2. *For every nilpotent sub-algebra* $\mathfrak{N}$ *of the total matric algebra* $\mathfrak{F}_n$ *of degree* n *over* $\mathfrak{F}$ *there is an automorphism* T *of* $\mathfrak{F}_n/\mathfrak{F}$ *carrying* $\mathfrak{N}$ *into* $\mathfrak{N}^T \leqq \mathfrak{F}^{(n)}$.

Proof. The algebra $\mathfrak{N}$ defines an algebra $\mathfrak{L}$, isomorphic to $\mathfrak{N}$, of linear transformations on the n-dimensional vector space V over $\mathfrak{F}$, and $\mathfrak{N}$ represents $\mathfrak{L}$ relative to the basis $e_1, \cdots, e_n$ where e_i is the vector whose i-th coordinate is 1 and whose other coordinates are 0. Thus if $N = (\nu_{ij}) \varepsilon \mathfrak{N}$, $N \rightarrow L$ where $e_i{}^L = \Sigma \nu_{ij} e_j$. The images v^L $(v \varepsilon V, L \varepsilon \mathfrak{L}^i)$ span a vector space $V\mathfrak{L}^i$ and V $(= V\mathfrak{L}^0) > V\mathfrak{L} > \cdots > V\mathfrak{L}^{\nu-1} > V\mathfrak{L}^{\nu} = 0$ where ν is the index of nilpotency of $\mathfrak{L}$. For, the fact that $\mathfrak{L}^{\nu-1} \neq 0$ implies that $V\mathfrak{L}^{\nu-1} \neq 0$. If $V\mathfrak{L}^i = V\mathfrak{L}^{i+1}$ for $i < \nu$, it follows that $V\mathfrak{L}^i = V\mathfrak{L}^{i+1} = \cdots = V\mathfrak{L}^{\nu} = 0$, a contradiction. Now we choose representatives of the classes constituting any basis of $V\mathfrak{L}^{\nu-1} - V\mathfrak{L}^{\nu}$, then of $V\mathfrak{L}^{\nu-2} - V\mathfrak{L}^{\nu-1}$, etc., and finally of $V - V\mathfrak{L}$, and number these quantities $v_1, v_2, \cdots, v_n$ in the order of choice. The quantities $v_1, \cdots, v_n$ so obtained form a basis of V. Relative to this basis each linear transformation T is represented by a matrix (α_{ij}) defined by the equation $v_i{}^T = \Sigma \alpha_{ij} v_j$ and if T is in $\mathfrak{L}$, the construction shows that $\alpha_{ij} = 0$ for $j \geqq i$. Thus $\mathfrak{F}^{(n)}$ contains the algebra $\mathfrak{N}_0$ of the (α_{ij}). Since $\mathfrak{N}_0$ and $\mathfrak{N}$ represent $\mathfrak{L}$ relative to different bases of V, it is known that $\mathfrak{N}_0 = P^{-1}\mathfrak{N}P$ where P is a non-singular matrix of $\mathfrak{F}_n$.

This result characterizes $\mathfrak{F}^{(n)}$ as a maximal algebra of nilpotent[4] linear transformations on the n-dimensional vector space over $\mathfrak{F}$. There is no such algebra of larger dimension and any other of the same dimension is isomorphic to $\mathfrak{F}^{(n)}$.

---

[2] A basis $u_1, \cdots, u_{n-1}$ such that each product $u_i u_j$ is a linear combination of those u_k with k greater than i and j. Cf. L. E. Dickson, *Algebras and their Arithmetics*, Chicago, 1923, pp. 175 f.

[3] Theorem 1 is clearly a corollary of Theorem 2. The separate proof of Theorem 1 is given both for its simplicity and for the notations and preliminary properties which it introduces.

[4] For an algebra $\mathfrak{A}$, nilpotency of all the quantities of $\mathfrak{A}$ is equivalent to the nilpotency of $\mathfrak{A}$.

2. Monic and inner automorphisms. If $\mathfrak{A}$ is an arbitrary (associative) ring, an automorphism T of $\mathfrak{A}$ is called "monic" in case $x^T - x$ lies in $\mathfrak{A}^{r+1}$ whenever x lies in $\mathfrak{A}^r$. This concept is trivial if $\mathfrak{A}$ possesses a unity but otherwise the concept is significant.

Let T and U be monic and consider the equation $x^{TU} - x = (x^{TU} - x^T) + (x^T - x)$. Since x^T lies in $\mathfrak{A}^r$ if x does, both quantities in parentheses lie in $\mathfrak{A}^{r+1}$, so that TU is monic. Also

$$x^T - x = y \varepsilon \mathfrak{A}^{r+1}, \qquad x^{T^{-1}} - x = - y^{T^{-1}} \varepsilon \mathfrak{A}^{r+1},$$

so that T^{-1} is monic. Hence the monic automorphisms form a group, $\mathfrak{M}$. This group is always a normal subgroup of the automorphism group $\mathfrak{G}$ of $\mathfrak{A}$. For, let us take $M \varepsilon \mathfrak{M}$, $G \varepsilon \mathfrak{G}$, $x \varepsilon \mathfrak{A}^r$. Then $x^G \varepsilon \mathfrak{A}^r$, $x^{GM} - x^G = y \varepsilon \mathfrak{A}^{r+1}$,

$$x^{GMG^{-1}} - x = y^{G^{-1}} \varepsilon \mathfrak{A}^{r+1}.$$

Thus GMG^{-1} is monic.

THEOREM 3. *The monic automorphisms of any associative ring*[5] $\mathfrak{A}$ *form a normal subgroup* $\mathfrak{M}$ *of the automorphism group of* $\mathfrak{A}$. *If* $\mathfrak{A}$ *is nilpotent, the group* $\mathfrak{M}$ *is nilpotent.*

To prove the latter statement, let $\mathfrak{M}_i$ be the subset of automorphisms $M \varepsilon \mathfrak{M}$ such that $x^M - x = x_i \varepsilon \mathfrak{A}^i$. It is easily verified that $\mathfrak{M}_i$ is a subgroup and that $\mathfrak{M} = \mathfrak{M}_2 \geqq \mathfrak{M}_3 \geqq \cdots \geqq \mathfrak{M}_\nu = I$, where $\mathfrak{A}^{\nu-1} > \mathfrak{A}^\nu = 0$. Then $\mathfrak{M}$ is nilpotent[6] if each commutator $(\mathfrak{M}, \mathfrak{M}_i)$ is contained in $\mathfrak{M}_{i+1}$. Using the convention that a subscript j on a quantity of $\mathfrak{A}$ indicates that the quantity belongs to $\mathfrak{A}^j$, we have for each $L \varepsilon \mathfrak{M}_i$ and each $M \varepsilon \mathfrak{M}$:

$$x^{MLM^{-1}} = (x + x_2)^{LM^{-1}} = (x + x_2 + x_i)^{M^{-1}} = x + x_i^{M^{-1}} = x + x_i + x_{i+1}$$
$$= (x + x_2 + x_i) + (x_{i+1} - x_2),$$
$$x^{MLM^{-1}L^{-1}} = (x + x_2) + (x_{i+1} - x_2)^{L^{-1}} = x + x_2 + x_{i+1} - x_2 + y_{i+1} = x + z_{i+1}.$$

This completes the proof.

Inner automorphisms may be defined in any ring $\mathfrak{A}$. An element $a \varepsilon \mathfrak{A}$ is called quasi-regular if there is a $b \varepsilon \mathfrak{A}$ such that $a + b + ab = a + b + ba = 0$. Then an "inner" automorphism J of $\mathfrak{A}$ is defined by the formula

$$x^J = x + bx + xa + bxa. \tag{2}$$

Proof that (2) gives an automorphism is obtained by considering a ring $\mathfrak{B}$

---

[5] This theorem and its proof remain valid if $\mathfrak{A}$ is an algebra over $\mathfrak{F}$ and the automorphisms considered are over $\mathfrak{F}$.

[6] H. Zassenhaus, *The Theory of Groups*, New York, 1949, pp. 125 f.

containing $\mathfrak{A}$ and a unity element 1. Then $1+a$ is regular with $1+b$ as its inverse in $\mathfrak{B}$. The inner automorphism $y \to (1+b)y(1+a) = y^J$ carries each $x \varepsilon \mathfrak{A}$ into a quantity $x^J \varepsilon \mathfrak{A}$ given by (2). These images x^J vary over every quantity $u \varepsilon \mathfrak{A}$ since $x = (1+a)u(1+b)$ is in $\mathfrak{A}$ and $x^J = u$. Thus J induces an automorphism in $\mathfrak{A}$.

The totality of automorphisms $J = J_a$ given by (2) as a varies over the quasi-regular elements of $\mathfrak{A}$ form a group $\mathfrak{J}$ and, from (2), it is evident that each J is monic, $\mathfrak{J} \leqq \mathfrak{M}$. If G is any automorphism of $\mathfrak{A}$,

$$x^{G^{-1}JG} = x + b^G x + x a^G + b^G x a^G$$

which is the same as (2) with a^G and b^G in place of a and b. But a^G is quasi-regular with b^G as its quasi-inverse, so that $G^{-1}JG$ is again an inner automorphism. In summary: The inner automorphisms of any associative ring $\mathfrak{A}$ are all monic and form a normal subgroup of the automorphism group of $\mathfrak{A}$.

3. Certain characteristic ideals. A subalgebra $\mathfrak{C}$ of an algebra $\mathfrak{A}$ over $\mathfrak{F}$ is called "characteristic" if every automorphism T of $\mathfrak{A}$ over $\mathfrak{F}$ induces an automorphism in $\mathfrak{C}$. The powers $\mathfrak{A}^k$ are characteristic ideals. The knowledge that x lies in a particular characteristic ideal of $\mathfrak{A}$ is a clue to the possible quantities x^T. We shall begin to track down the automorphisms T of $\mathfrak{A} = \mathfrak{F}^{(n)}$ by finding for each matric unit e_{ij} in $\mathfrak{A}$ certain characteristic ideals containing e_{ij}.

Define $\mathfrak{C}_l$ to be the totality of matrices in $\mathfrak{F}^{(n)}$ whose columns beyond the l-th are zero. Then $\mathfrak{C}_l$ is an ideal of $\mathfrak{F}^{(n)}$. Likewise, an ideal is given by the set $\mathfrak{R}_k$ of all matrices in $\mathfrak{F}^{(n)}$ whose first $k-1$ rows are zero. We shall see[7] that $\mathfrak{C}_l$ and $\mathfrak{R}_k$ are characteristic.

The algebra $\mathfrak{A}^l$ consists (see footnote 1) of all $n \times n$ matrices with zeros everywhere except for an area in the lower left hand corner which may be described as a right triangle with both legs having length $n-l$ units. The elements in this area are completely arbitrary. It follows quickly that $\mathfrak{C}_l$ is precisely the totality of quantities $c \varepsilon \mathfrak{F}^{(n)} = \mathfrak{A}$ such that $c\mathfrak{A}^l = 0$, that is, $\mathfrak{C}_l$ is the left annihilator of $\mathfrak{A}^l$ in $\mathfrak{A}$. Then for each $c \varepsilon \mathfrak{C}_l$, $x \varepsilon \mathfrak{A}^l$, and automorphism T of $\mathfrak{A}$, we have $cx = 0 = (cx)^T = c^T x^T$, and since x^T varies over all of $\mathfrak{A}^l$ as x does, $c^T \mathfrak{A}^l = 0$, so $c^T \varepsilon \mathfrak{C}_l$. Thus $\mathfrak{C}_l$ is a characteristic ideal of $\mathfrak{A}$.

The ideal $\mathfrak{R}_k$ can be described as the totality of quantities $r \varepsilon \mathfrak{A} = \mathfrak{F}^{(n)}$ such that $\mathfrak{C}_{k-1} r = 0$ (the right annihilator of $\mathfrak{C}_{k-1}$ in $\mathfrak{A}$). Since $\mathfrak{C}_{k-1}$ is characteristic, the final argument above may be paraphrased to show that $\mathfrak{R}_k$ is characteristic.

---

[7] We are indebted to the referee for simplifications in the proofs of Theorems 4 and 7.

THEOREM 4. *The ideals* $\mathfrak{C}_l$ *and* $\mathfrak{R}_k$ *of* $\mathfrak{F}^{(n)}$ *are characteristic. For each automorphism* T *of* $\mathfrak{F}^{(n)}$ *and each matric unit* $e_{kl} \varepsilon \mathfrak{F}^{(n)}$, $e_{kl}{}^T$ *lies in* $\mathfrak{C}_l \cap \mathfrak{R}_k$.

Thus, in particular,

$$(3) \qquad e_{k,k-1}{}^T = \sum_{i \geq k}^{j<k} \beta_{ij}{}^{(k)} e_{ij} \qquad (\beta_{ij}{}^{(k)} \varepsilon \mathfrak{F}).$$

The "corner" coefficient

$$\beta_{k,k-1}{}^{(k)} = \alpha_k$$

cannot be zero for otherwise $e_{k,k-1}{}^T$ would lie in the square of $\mathfrak{F}^{(n)}$ whence so would $e_{k,k-1}$, contrary to fact. Moreover, if T is monic, each $\alpha_k = 1$; the converse can be verified easily from the fact that $\mathfrak{F}^{(n)}$ is generated by the quantities $e_{k,k-1}$.

4. Diagonal automorphisms. The set $\mathfrak{T}_n$ of all matrices in $\mathfrak{F}_n$ with zeros above the diagonal is an algebra whose radical is $\mathfrak{F}^{(n)}$. The radical of any algebra is characteristic and we shall be interested in the automorphisms induced in $\mathfrak{F}^{(n)}$ by certain inner automorphisms of $\mathfrak{T}_n$.

Each diagonal non-singular matrix, $d = \Sigma \delta_i e_{ii}$ $(\delta_i \varepsilon \mathfrak{F})$, determines an automorphism $x^D = dxd^{-1}$ of $\mathfrak{T}_n$, and the automorphism which D induces in $\mathfrak{F}^{(n)}$ is called a "diagonal" automorphism of $\mathfrak{F}^{(n)}$. We have

$$(4) \qquad e_{k,k-1}{}^D = \alpha_k e_{k,k-1},$$

and

$$(5) \qquad e_{kl}{}^D = \alpha_{kl} e_{kl}, \qquad \alpha_{kl} = \alpha_k \alpha_{k-1} \cdots \alpha_{l+1},$$

where the α_k are non-zero scalars. Conversely, arbitrary non-zero scalars $\alpha_2, \cdots, \alpha_n$ and equations (4) and (5) determine a diagonal automorphism given by $d = e_{11} + \alpha_2 e_{22} + \cdots + (\alpha_2 \cdots \alpha_n) e_{nn}$. Each diagonal D is obtained precisely once in the form $x^D = dxd^{-1}$ where $d = e_{11} + \delta_2 e_{22} + \cdots + \delta_n e_{nn}$, and, consequently, the group $\mathfrak{D}$ of all diagonal automorphisms of $\mathfrak{F}^{(n)}$ is isomorphic to the multiplicative group of non-singular diagonal matrices in $\mathfrak{F}_{n-1}$.

Next we find that from each automorphism T a unique diagonal factor D may be removed so that $D^{-1}T$ is monic.

THEOREM 5. *The automorphism group of* $\mathfrak{F}^{(n)}$ *has the structure* $\mathfrak{G} = \mathfrak{M}\mathfrak{D} = \mathfrak{D}\mathfrak{M}$ *where* $\mathfrak{D}$ *is the group of diagonal automorphisms,* $\mathfrak{M}$ *the group of monic automorphisms, and* $\mathfrak{D} \cap \mathfrak{M} = I$.

In (3) and the remarks after (3) it appeared that certain "corner coefficients" $\alpha_k = \beta_{k,k-1}{}^{(k)}$ cannot be zero. By multiplying equations like (3) for the quantities $e_{k,k-1}{}^T, \cdots, e_{l+1,l}{}^T$ we find that

$$e_{kl}{}^T = \sum_{i \geqq k}^{j \leqq l} \alpha_{ij} e_{ij}$$

with corner coefficient $\alpha_{kl} = \alpha_k \alpha_{k-1} \cdots \alpha_{l+1}$. The quantities α_k, α_{kl} thus fulfill the conditions in (5) so that the correspondence $e_{kl} \rightarrow \alpha_{kl} e_{kl}$ generates a diagonal automorphism D. Then $D^{-1}T = T_1$ and $TD^{-1} = T_2$ are automorphisms whose corner coefficients are all unity and this implies that T_1 and T_2 are monic.

That $\mathfrak{D} \cap \mathfrak{M} = I$ is evident from (4) and (5). Thus in any expression $T = DM = M_1 D_1$ the factors are unique and, actually, $D_1 = D$ since $\mathfrak{M}$ is normal in $\mathfrak{G}$.

The idea of inner automorphisms of Section **2** may be applied to $\mathfrak{A} = \mathfrak{F}^{(n)}$. Here every $a \in \mathfrak{A}$ is quasi-regular with quasi-inverse $b = -a + a^2 - \cdots + (-1)^t a^t$ $(a^{t+1} = 0)$. With the aid of the unity in $\mathfrak{T}_n$ we may write $g = 1 + a = 1 + \Sigma \alpha_{ij} e_{ij} = g_2 g_3 \cdots g_n$, where

$$g_i = 1 + \sum_{j=1}^{i-1} \alpha_{ij} e_{ij} = 1 + a_i. \tag{6}$$

Then $g_i^{-1} = 1 - a_i$ and

$$x^J = g^{-1} x g = g_n{}^{-1} \cdots g_2{}^{-1} x g_2 \cdots g_n, \qquad J = J_a = J_{a_2} \cdots J_{a_n}. \tag{7}$$

These properties will be needed soon.

5. Nil automorphisms.

A quantity $w \in \mathfrak{A}$ is an absolute right (left) divisor of zero if $xu = 0$ $(ux = 0)$ for every $x \in \mathfrak{A}$. For $\mathfrak{A} = \mathfrak{F}^{(n)}$ these are the quantities $e_{n1}, e_{n2}, \cdots, e_{n,n-1}$ $(e_{21}, e_{31}, \cdots, e_{n1})$ and their linear combinations.

An automorphism N of $\mathfrak{A}$ is called "nil" if $u^N = u$ for every absolute right or left divisor of zero. It is clear that the totality of nil automorphisms of any ring form a group $\mathfrak{N}$ which is actually a normal subgroup of the full automorphism group.

THEOREM 6. *All nil automorphisms of* $\mathfrak{F}^{(n)}$ *are generated by*

$$e_{k,k-1}{}^N = e_{k,k-1} + \gamma_k e_{n1} \qquad (k = 2, \cdots, n) \tag{8}$$

where $\gamma_2 = \gamma_n = 0$ *and the remaining* γ_k *are arbitrary scalars.*

From (8) it follows that $e_{kj}{}^N = e_{kj}$ for $j < k - 1$, whence $e_{k1}{}^N = e_{k1}$ and $e_{nj}{}^N = e_{nj}$ so that N is nil. Conversely, let N be nil. By Theorem 5 $N = DM$,

$$e_{k,k-1}{}^D = \alpha_k e_{k,k-1}, \qquad e_{k,k-1}{}^{DM} = \alpha_k e_{k,k-1} + c_k$$

with $c_k \varepsilon \mathfrak{A}^2$. Then $\alpha_2 = 1$ and $c_2 = 0$. If $k > 2$,

$$e_{k1} = e_{k1}{}^N = (e_{k,k-1}e_{k-1,1})^N = e_{k,k-1}{}^{DM}e_{k-1,1} = (\alpha_k e_{k,k-1} + c_k)e_{k-1,1}$$
$$= \alpha_k e_{k1} + c_k e_{k-1,1}$$

so $\alpha_k = 1$ and $c_k e_{k-1,1} = 0$. By the first of these facts, $D = I$ and N is monic. By the second the $(k-1)$-st column of the matrix c_k is 0. For $1 < j < k-1$,

$$0 = (e_{k,k-1}e_{j1})^N = (e_{k,k-1} + c_k)e_{j1} = c_k e_{j1}$$

so that the j-th column of c_k is 0 for all $j > 1$. For $n > j > k$,

$$0 = (e_{nj}e_{k,k-1})^N = e_{nj}{}^N e_{k,k-1}{}^N = e_{nj}(e_{k,k-1} + c_k) = e_{nj}c_k$$

so the j-th row of c_k is 0, $j = k+1, \cdots, n-1$. Also

$$(e_{nk}e_{k,k-1})^N = e_{n,k-1}{}^N = e_{n,k-1} = e_{nk}(e_{k,k-1} + c_k) = e_{n,k-1} + e_{nk}c_k,\ e_{nk}c_k = 0$$

so that the k-th row of c_k is also 0, $c_k = \gamma_k e_{n1}$, and this gives (8).

THEOREM 7. *The group of monic automorphisms of $\mathfrak{F}^{(n)}$ has the structure $\mathfrak{M} = \mathfrak{N} \times \mathfrak{J}$, where $\mathfrak{N}$ is the group of nil automorphisms and $\mathfrak{J}$ the group of inner automorphisms.*

To prove that $\mathfrak{M} = \mathfrak{N}\mathfrak{J}$ it suffices to find for each $M \varepsilon \mathfrak{M}$ an automorphism $J \varepsilon \mathfrak{J}$ such that $MJ = N \varepsilon \mathfrak{N}$, whence $M = NJ^{-1}$. We have

$$e_{k1}{}^M = e_{k1} + \sum_{p>k} \alpha_{pk}e_{p1} \qquad (k = 2, \cdots, n-1).$$

The scalars α_{pk} are thus defined for $n \geqq p > k > 1$, and

$$e_{k1}{}^M = e_{k1} + \sum_{p>q>1} \alpha_{pq}e_{pq}e_{k1} = (1+a)e_{k1} = (1+a)e_{k1}(1+a)^{-1},$$
$$a = \sum_{p>q>1} \alpha_{pq}e_{pq}.$$

The inner automorphism $x^{J_1} = g^{-1}xg$, $g = 1 + a$, has the property that $M_1 = MJ_1$ fixes each element e_{k1}. Since M_1 is monic, $e_{nk}{}^{M_1} = e_{nk} + \sum_{q<k} \beta_{kq}e_{nq}$, and for all p such that $1 < p < k$ we have $e_{nk}e_{p1} = 0 = (e_{nk}e_{p1})^{M_1} = e_{nk}{}^{M_1}e_{p1} = \sum_{q<k} \beta_{kq}e_{nq}e_{p1} = \beta_{kp}e_{n1}$. It follows that the coefficients β_{kp} are zero except possibly for $\beta_{k1} = \beta_k$ $(k = 2, \cdots, n-1)$ so

$$e_{nk}{}^{M_1} = e_{nk} + \beta_k e_{n1} = e_{nk} + e_{nk}\sum_{j=2}^{n-1} \beta_j e_{j1}$$
$$= e_{nk}(1+b), \qquad b = \sum_{j=2}^{n-1} \beta_j e_{j1}.$$

Then the inner automorphism $x^{J_2} = (1+b)x(1+b)^{-1}$ has the properties

$$e_{nk}{}^{M_1J_2} = e_{nk}, \qquad (k = 1, \cdots, n-1),$$
$$e_{k1}{}^{M_1J_2} = e_{k1}{}^{J_2} = (1+b)e_{k1}(1+c) = e_{k1}.$$

Thus $M_1J_2 = MJ_1J_2$ is nil. This proves $\mathfrak{M} = \mathfrak{N}\mathfrak{J}$.

Since $\mathfrak{N}$ and $\mathfrak{J}$ are normal subgroups of $\mathfrak{M}$, the product $\mathfrak{N}\mathfrak{J}$ is the direct product $\mathfrak{N} \times \mathfrak{J}$ if and only if the intersection $\mathfrak{N} \cap \mathfrak{J}$ is the identity. Let J be in this intersection: $x^J = (1+a)^{-1}x(1+a)$. Since J also lies in $\mathfrak{N}$, $e_{k1}{}^J = e_{k1}$, $(1+a)e_{k1} = e_{k1}(1+a)$ whence $ae_{k1} = e_{k1}a = 0$. The matrix a thus has all columns zero except possibly its first column. Similarly, J fixes each e_{nk}, $0 = ae_{nk} = e_{nk}a$, and a has only zero rows except perhaps for row n. Thus $a = \alpha e_{n1}$, $\alpha \varepsilon \mathfrak{F}$,

$$(1+a)^{-1} = 1 - \alpha e_{n1}, \qquad x^J = (1 - \alpha e_{n1})x(1 + \alpha e_{n1}) = x.$$

This proves $J = I$ and completes the proof of the theorem.

6. The automorphism group of $\mathfrak{T}_n$. Each diagonal automorphism $x^D = d^{-1}xd$ and each inner automorphism $x^J = (1+a)^{-1}x(1+a)$ of $\mathfrak{F}^{(n)}$ may be extended to $\mathfrak{T}_n$ over $\mathfrak{F}$ simply by letting x vary over $\mathfrak{T}_n$. The theorem below implies that the automorphisms DJ of $\mathfrak{F}^{(n)}$ are the only ones extendable to $\mathfrak{T}_n$ and it determines all extensions of these.

THEOREM 8. *The group $\mathfrak{E}$ of automorphisms of $\mathfrak{T}_n/\mathfrak{F}$ inducing the identity on $\mathfrak{F}^{(n)}$ is given by $x^E = g^{-1}xg = (1 - \alpha e_{n1})x(1 + \alpha e_{n1})$ where α is an arbitrary scalar. All automorphisms of $\mathfrak{T}_n/\mathfrak{F}$ are inner and the group they induce in $\mathfrak{F}^{(n)}$ is $\mathfrak{D}\mathfrak{J}$.*

Since each automorphism T of $\mathfrak{T}_n/\mathfrak{F}$ maps the e_{ii} into a set of mutually orthogonal idempotents whose sum is 1, it can be shown that

$$e_{kk}{}^T = e_{j_kj_k} + b_k \tag{9}$$

where b_k is in $\mathfrak{F}^{(n)}$ and $k \rightarrow j_k$ is a permutation of $1, 2, \cdots, n$; and

$$e_{j_kj_k}b_k + b_ke_{j_kj_k} + b_k{}^2 = b_k, \tag{10}$$

$$\sum_{k=1}^{n} b_k = 0. \tag{11}$$

To show that the nil automorphisms N are not extendable to $\mathfrak{T}_n$ unless $N = I$, let T induce N on $\mathfrak{A} = \mathfrak{F}^{(n)}$. Then by Theorem 6,

$$\begin{aligned}(e_{kk}e_{k,k-1})^T = e_{k,k-1} + \alpha_ke_{n1} &= (e_{j_kj_k} + b_k)(e_{k,k-1} + \alpha_ke_{n1})\\ &= e_{j_kj_k}e_{k,k-1} + c_k \qquad (c_k \varepsilon \mathfrak{A}^2, k = 2, \cdots, n),\end{aligned}$$

so $e_{j_kj_k}e_{k,k-1} = e_{k,k-1}$, $j_k = k$ $(k = 2, \cdots, n)$ whence also $j_1 = 1$ and

$$e_{kk}^T = e_{kk} + b_k \qquad (k = 1, \cdots, n).$$

Also

$$\begin{aligned}(e_{k,k-1}e_{k-1,k-1})^T = e_{k,k-1} + \alpha_k e_{n1} &= (e_{k,k-1} + \alpha_k e_{n1})(e_{k-1,k-1} + b_{k-1}) \\ &= e_{k,k-1} + (e_{k,k-1}b_{k-1} + \delta\alpha_k e_{n1}) \qquad (k = 2, \cdots, n),\end{aligned}$$

where $\delta = 0$ unless $k - 1 = 1$. Then the quantity in parenthesis must equal $\alpha_k e_{n1}$. If T is not the identity on $\mathfrak{J}^{(n)}$, there is a value of k such that $n > k > 2$ and $\alpha_k \neq 0$ (Theorem 6). Choosing this k we find $\delta = 0$, $e_{k,k-1}b_{k-1} = \alpha_k e_{n1} \neq 0$, which is impossible since the n-th row of the left side is 0. This proves that nil automorphisms $\neq I$ are not extendable, so the only extendable automorphisms are those of $\mathfrak{D}\mathfrak{J}$.

If T is any automorphism of $\mathfrak{T}_n$, it induces an automorphism DJ in $\mathfrak{J}^{(n)}$, and the latter is clearly extendable to an inner automorphism S of $\mathfrak{T}_n$ such that $TS^{-1} = E$ is the identity on $\mathfrak{J}^{(n)}$. It remains only to show that each automorphism E inducing the identity in $\mathfrak{J}^{(n)}$ is inner and obeys the formula of Theorem 8. Then $T = ES$ will be inner.

For $i < k = 2, \cdots, n$ we have, by (9), $(e_{kk}e_{ki})^E = e_{ki} = (e_{j_kj_k} + b_k)e_{ki}$, whence $j_k = k$ and $b_k \in \mathfrak{J}^{(n)}$ has k-th column zero. Then $b_ke_{kk} = 0$ and (10) yields

$$b_k = e_{kk}b_k + b_k^2 \qquad (k = 2, \cdots, n). \tag{12}$$

On squaring both sides we find

$$b_k^2 = e_{kk}b_k^3 + b_k^4 \tag{13}$$

which implies that b_k^2 lies in $\mathfrak{A}^3$, $\mathfrak{A} = \mathfrak{J}^{(n)}$. When this fact is employed on the right side of (13) it turns out that b_k^2 lies in $\mathfrak{A}^4$ and, similarily, b_k^2 lies in $\mathfrak{A}^t$, $t = 3, 4, \cdots$. But then $b_k^2 = 0$ and (12) reduces to $b_k = e_{kk}b_k$, that is, b_k is zero except possibly for its k-th row $(k = 2, \cdots, n)$. A parallel proof shows that b_k is 0 except possibly for its k-th column $(k = 1, \cdots, n - 1)$. It follows that $b_2 = \cdots = b_{n-1} = 0$ and, from (11), $b_1 = -b_n$. But b_1 has only a first column, b_n only an n-th row; so $b_n = \alpha e_{n1}$, $b_1 = -\alpha e_{n1}$. Thus $e_{ij}^E = e_{ij}$ except for $e_{11}^E = e_{11} - \alpha e_{n1}$, $e_{nn}^E = e_{nn} + \alpha e_{n1}$, and this is the automorphism E given in Theorem 8.

7. Ideals. A set of r matric units $f_i = e_{k_il_i} \in \mathfrak{J}^{(n)}$ will be called a "set of corners" if $r < n$ and $k_1 < k_2 < \cdots < k_r$, $l_1 < l_2 < \cdots < l_r$. Each corner may be represented by the point in an $n \times n$ matrix where its only non-zero coordinate occurs, and these points determine a "staircase" as follows. Regard the matrix as a square array of n^2 points, and draw a line in row k_1 extending from the left boundary of the square to the corner $f_1 = e_{k_1 l_1}$. Continue the line from this corner down to row k_2, then to the right until it reaches f_2. In general, if $i < r$, the line goes from corner f_i down to row k_{i+1}, then right to corner f_{i+1}. From f_r the line simply goes down to row n; or the line stops at f_r if f_r lies in row n. The open polygon so obtained has the appearance of a staircase in which each corner, except possibly f_r, is at the top of a riser. We shall refer to this open polygon as the "staircase" determined by the given set of corners. A matric unit e_{ij} is an "interior quantity" for the set of corners if its representative point lies below the staircase or on it but not at a corner.

THEOREM 9. *Let $f_1, \cdots, f_r$ be any set of corners and $u_1 \cdots, u_s$ be linear combinations of these corners with scalar coefficients such that each f_i occurs in at least one u_j with non-zero coefficient. Then $u_1, \cdots, u_s$ and the interior quantities for the set of corners span an ideal of $\mathfrak{F}^{(n)}$ and, conversely, every ideal is obtained in this way.*

Proof that the space spanned is an ideal is routine. In proving that every ideal $\mathfrak{B}$ is obtained in the manner described we first find the corners. Among all the e_{ij} occurring in the expressions of the quantities of $\mathfrak{B}$ in terms of matric units, let $i = k_1$ denote the minimum value of i. Among all the $e_{k_1 j}$ occurring let $j = l_1$ be maximum. This determines f_1. Among all e_{ij} occurring with $i > k_1$ and $j > l_1$ (if any) let $i = k_2$ denote a minimum value and, for this i, let $j = l_2$ be maximum. This determines f_2 and a continuation yields all the corners $f_1, \cdots, f_r$, $r < n$. The non-zero coordinates of all matrices in $\mathfrak{B}$ occur on or under the staircase and it will be shown that $\mathfrak{B}$ contains all the interior quantities.

For each f_i there is a quantity u_i in $\mathfrak{B}$ containing f_i with coefficient unity. Let $f_i = f = e_{kl}$ and $u_i = u$. If $f = e_{n1}$, this must be the only corner and $\mathfrak{B}$ has the form required, so we assume that e_{n1} is not a corner. If $l = 1$, $\mathfrak{B}$ contains $e_{jk}u = e_{j1}$ $(j = k + 1, \cdots, n)$ and if $k = n$, $\mathfrak{B}$ contains $ue_{lm} = e_{nm}$ $(m = 1, \cdots, l - 1)$. If $l > 1$ and $k < n$, $\mathfrak{B}$ contains every $e_{jk}ue_{lm} = e_{jm}$ $(j = k + 1, \cdots, n; m = 1, \cdots, l - 1)$, that is, every e_{jm} lying below and to the left of $f = e_{kl}$. Since this is true for every f not at an end of the staircase, $\mathfrak{B}$ contains every interior quantity under the staircase. If any of the u_i involve interior quantities under the staircase, these may be subtracted off, so it will be assumed henceforth that the u_i are linear combinations of the corners and possibly other matric units lying on the staircase.

If u and $f = e_{kl}$ are as above, $\mathfrak{B}$ contains $e_{jk}u = e_{jl} + v$ where v is a linear combination of interior quantities, hence in $\mathfrak{B}$, so $\mathfrak{B}$ contains e_{jl} for all $j > k$. Likewise, $\mathfrak{B}$ contains all e_{km}, $m < l$, that is, $\mathfrak{B}$ contains every staircase quantity to the left of or below any corner. By use of these quantities the u_i may be altered so that they involve only corners.

If $x \varepsilon \mathfrak{B}$, subtraction of interior quantities leaves a quantity $y = \Sigma \alpha_i f_i \varepsilon \mathfrak{B}$. If y is not a linear combination of the u_i, it will be taken as an additional quantity u. This process stops after a finite number of repetitions since all the quantities y and u lie in the r-dimensional vector space spanned by the corners. The u_i now at hand, together with the interior quantities, span $\mathfrak{B}$. This completes the proof.

8. Characteristic ideals. A subalgebra $\mathfrak{B}$ of $\mathfrak{A}$ is called "characteristic" if $\mathfrak{B}^T = \mathfrak{B}$ for every automorphism T of $\mathfrak{A}$. If $\mathfrak{B}$ is an ideal of $\mathfrak{F}^{(n)}$ and M is a monic automorphism, M carries each e_{ij} in $\mathfrak{B}$ or u_i (Theorem 9) in $\mathfrak{B}$ into the same quantity plus an interior quantity of $\mathfrak{B}$, so $\mathfrak{B}^M = \mathfrak{B}$. Therefore $\mathfrak{B}$ is characteristic if and only if $\mathfrak{B}^D = \mathfrak{B}$ for each diagonal automorphism D. This condition is satisfied if the corners are contained in $\mathfrak{B}$ since D merely multiplies each matric unit by a scalar. By the next theorem this is essentially the only case in which $\mathfrak{B}$ is a characteristic ideal.

THEOREM 10. *If $\mathfrak{F}$ contains only two elements, every ideal of $\mathfrak{F}^{(n)}$ is characteristic. Otherwise, an ideal $\mathfrak{B}$ of $\mathfrak{F}^{(n)}$ is characteristic if and only if $\mathfrak{B}$ contains its corners.*

The proof of the first statement is the observation that every diagonal automorphism is the identity.

LEMMA 1. *If $f_1, \cdots, f_r$ are a set of corners and $\phi_1, \cdots, \phi_r$ are arbitrary non-zero scalars, there is a diagonal automorphism D such that $f_i^D = \phi_i f_i$ $(i = 1, \cdots, r)$.*

First we construct $D = D_a$ such that

$$f_a^D = \phi_a f_a, \qquad f_i^D = f_i \qquad (i \neq a), \tag{14}$$

by finding a suitable diagonal matrix $d = \delta_1 e_{11} + \cdots + \delta_n e_{nn}$ to use for $x^D = dxd^{-1}$. Let $f_a = f = e_{kl}$ so $l < k$ and $f^D = \delta_k \delta_l^{-1} f$. Then the definitions $\delta_k = \phi_a$, $\delta_j = 1$ for $j < k$, yield (14) for $i = 1, \cdots, a$. Let $f_j = e_{k_j l_j}$ for $a \leqq j < r$ and assume that $\delta_1, \cdots, \delta_{k_j}$ have been defined so that (14) holds for $i = 1, \cdots, j$ and that for $h > k_j$, δ_h is still undefined. If $f_{j+1} = e_{pq}$ then $f_{j+1}{}^D = \delta_p \delta_q{}^{-1} f_{j+1}$ and δ_p is still undefined. If δ_q is among the δ's already defined, we take $\delta_p = \delta_q$, and $\delta_h = 1$ for $k_j < h < p$. If δ_q is undefined, we take $\delta_h = 1$ for $k_j < h \leqq p$. Then $f_{j+1}{}^D = f_{j+1}$, whence the induction is completed. The D of the lemma is given by $D = D_1 D_2 \cdots D_r$.

Now let $\mathfrak{F}$ contain more than two elements. The sufficiency has been noted above, so it remains to show that a characteristic ideal $\mathfrak{B}$ contains its corners. The u_i of Theorem 9 may be taken to be linearly independent, so $s \leqq r$ and the equations

$$u_i = \sum_{j=1}^{r} \beta_{ij} f_j \qquad (i = 1, \cdots, s)$$

define an $s \times r$ matrix $\mathrm{B} = (\beta_{ij})$ of rank s. Since $\mathfrak{B}$ is characteristic it contains

$$w_i{}^D = \sum_{j=1}^{r} \beta_{ij} \phi_j f_j \qquad (i = 1, \cdots, s) \tag{15}$$

for all non-zero $\phi_j \in \mathfrak{F}$ (Lemma 1). But if $u_i{}^D$ is in $\mathfrak{B}$ it must be of the form $\Sigma \xi_{ih} u_h$, which implies that

$$\sum_j \beta_{ij} \phi_j f_j = \sum_{h,j} \xi_{ih} \beta_{hj} f_j \qquad (i = 1, \cdots, s)$$

or, in matrix form,

$$\mathrm{B}\Phi = \Xi \mathrm{B}, \tag{16}$$

where $\Xi = (\xi_{ik})$ is an $s \times s$ matrix depending on the diagonal matrix Φ whose diagonal elements are $\phi_1, \cdots, \phi_r$. Since B is $s \times r$ of rank s it has a right inverse Δ, that is, an $r \times s$ matrix Δ such that $\mathrm{B}\Delta = I_s$, the s-rowed identity. Therefore $\Xi = \mathrm{B}\Phi\Delta$ and (16) implies that

$$\mathrm{B}\Phi\mathrm{H} = 0, \qquad \mathrm{H} = (\eta_{ij}) = I_r - \Delta\mathrm{B}. \tag{17}$$

Note that H is independent of Φ and that (17) is true for all choices of Φ. Taking $\phi_j = 1$ for $j \neq a$ we find that the element in the (i, l) place in (17) is

$$\beta_{ia} \phi_a \eta_{al} + \sum_{j \neq a} \beta_{ij} \eta_{jl} = 0 \tag{18}$$

for all choices of $\phi_a \neq 0$. Thus (18) is a linear equation in ϕ_a with more than one solution (note the hypothesis on F!)—a contradiction unless every $\beta_{ia}\eta_{al} = 0$. For each value of a there is a value of i such that $\beta_{ia} \neq 0$ (each f_a occurs in some u_i). Therefore, every $\eta_{al} = 0$, $\mathrm{H} = 0$, and $\Delta\mathrm{B} = I_r$. Since $\Delta\mathrm{B}$ cannot have higher rank than B, the matrix B must have rank r. Hence $s = r$. Since the space $(f_1, \cdots, f_r)$ of dimension r has a subspace $(u_1, \cdots, u_r)$ of the same dimension, the two coincide, whence all the f_i lie in $\mathfrak{B}$.

COROLLARY. *If* $\mathfrak{F}$ *is not of characteristic two, the number of characteristic ideals of* $\mathfrak{F}^{(n)}$ *is* $\frac{1}{n}\binom{2n}{n-1}$.

The proof of this fact, which was kindly supplied us by G. Szekeres, will be omitted here.

9. Elementary nilpotent algebras. An ideal with a single corner $f = e_{kl}$ is the intersection $\mathfrak{D}_{kl} = \mathfrak{R}_k \cap \mathfrak{C}_l$ of the characteristic ideals $\mathfrak{R}_k$ and $\mathfrak{C}_l$ defined in Section **3** and, therefore, every characteristic ideal is a join of such intersections $\mathfrak{D}_{kl}$. This is one of several facts lending interest to the ideals $\mathfrak{C}_l = \mathfrak{C}_l{}^{(n)}$ as a class of abstract algebras. It is unnecessary to study also the algebras $\mathfrak{R}_k$ since $k = n - l + 1$ for an appropriate l and $\mathfrak{R}_{n-l+1} \cong \mathfrak{C}_l$ under the correspondence $e_{n-j+1,n-i+1} \rightarrow e_{ij}$. The algebras $\mathfrak{D}_{kl}$ are of no structural interest since they are zero algebras.

Any algebra over $\mathfrak{F}$ isomorphic to $\mathfrak{C}_l{}^{(n)}$ will be called an elementary nilpotent algebra. Note that $\mathfrak{F}^{(n)} = \mathfrak{C}_{n-1}{}^{(n)}$ so that $\mathfrak{F}^{(n)}$ is elementary. By direct computation, the index of $\mathfrak{C}_l{}^{(n)}$ is found to be $l + 1$.

THEOREM 11. *Two elementary nilpotent algebras are isomorphic if and only if they have the same values of* l *and of* n. *The invariant* l *is equal to* $\nu - 1$ *where* ν *is the index of nilpotency.*

The sufficiency is evident. Suppose, conversely, that $\mathfrak{C}_k{}^{(m)} \cong \mathfrak{C}_l{}^{(n)}$. The common index of nilpotency of these algebras is $k + 1 = l + 1$, whence $k = l$. The common order is $l[n - (l+1)/2] = l[m - (l+1)/2]$, so $n = m$.

THEOREM 12. $\mathfrak{C}_l{}^{(n)}$ *is indecomposable unless* $l = 1$.

Since $\mathfrak{C}_l{}^{(n)}$ is a zero algebra, it is the direct sum of $n - 1$ ideals of order one. Now let l be greater than 1 and suppose that $\mathfrak{C}_l{}^{(n)} = \mathfrak{A} \oplus \mathfrak{B}$. Then one of $\mathfrak{A}$ and $\mathfrak{B}$, say $\mathfrak{A}$, contains an element of the form

$$a = e_{21} + \sum_{i \neq 2} \alpha_{ij} e_{ij} \qquad (\alpha_{ij} \in \mathfrak{F}) \tag{19}$$

whence $\mathfrak{A}$ contains every $e_{j2}a = e_{j1}$ $(j = 3, \cdots, n)$. If $\mathfrak{B}$ contained an element (19) it would also contain e_{n1}, contrary to the fact that $\mathfrak{A} \cap \mathfrak{B} = 0$. Hence every non-zero $b \in \mathfrak{B}$ has the form $b = \Sigma \beta_{ij} e_{ij}$ with $\beta_{21} = 0$. Then if all $\beta_{ij} = 0$ for $j > 1$, b would lie in $\mathfrak{A}$, $b = 0$, a contradiction implying that some $\beta_{pq} \neq 0$ for $q > 1$. Then $\mathfrak{B}$ contains $be_{q1} = \Sigma \beta_{iq} e_{i1} \neq 0$ and be_{q1} is a linear combination of $e_{31}, \cdots, e_{n1}$. Hence be_{q1} is in $\mathfrak{A} \cap \mathfrak{B}$. This contradiction shows that $\mathfrak{B}$ has no quantities $b \neq 0$.

The center of any nilpotent algebra $\mathfrak{N}$ of index ν contains the highest non-zero power $\mathfrak{N}^{\nu-1}$ of the algebra and the center is "minimal" if it equals this power. The center of $\mathfrak{C}_l^{(n)}$ (hence, also, of $\mathfrak{F}^{(n)}$) is minimal. A more inclusive result is given now.

THEOREM 13. *Let* $\mathfrak{A} = \mathfrak{C}_l^{(n)}$. *Then the difference algebras* $\mathfrak{A} - \mathfrak{A}^h$ $(h = 2, 3, \cdots, l+1)$ *have minimal centers* $\mathfrak{A}^{h-1} - \mathfrak{A}^h$.

First it should be noted that $\mathfrak{A}$ is generated by the matric units on its "right border," that is, by $e_{21}, e_{32}, \cdots, e_{l+1,l}, e_{l+2,l}, \cdots, e_{nl}$. Also the successive powers of $\mathfrak{A}$ are obtained by removing successive lines along the right border.

The theorem is true for $h = 2$ since $\mathfrak{A} - \mathfrak{A}^2$ is a zero algebrra. Assuming the result for $\mathfrak{A} - \mathfrak{A}^h$ $(h < l+1)$ we now let $cx - xc$ lie in $\mathfrak{A}^{h+1}$ for every x of $\mathfrak{A}$ and wish to prove c in $\mathfrak{A}^h$. Since $cx - xc$ is always in $\mathfrak{A}^h$, c must be in $\mathfrak{A}^{h-1}$, $c = c_0 + d$ where c_0 is in $\mathfrak{A}^h$ and

$$d = \sum_{j=1}^{l-h+1} \alpha_{h-1+j,j} e_{h-1+j,j} + \sum_{k=l+1}^{n} \alpha_{k,l-h+2} e_{k,l-h+2}. \tag{20}$$

It suffices to prove $d = 0$. From $d = c - c_0$ it follows that $dx - xd \,\varepsilon\, \mathfrak{A}^{h+1}$ for every x. Hence $\mathfrak{A}^{h+1}$ contains the quantities $a_i = de_{i,i-1} - e_{i,i-1}d$ $(i = 2, \cdots, l)$. The non-zero elements of the matrix $b_i = de_{i,i-1}$ lie in the $(i-1)$-st column and those of $c_i = e_{i,i-1}d$ lie to the left of this column. Then the fact that a_i lies in $\mathfrak{A}^{h+1}$ implies that b_i and c_i also lie in $\mathfrak{A}^{h+1}$. But $b_i = \alpha_{h-1+i,i} e_{h-1+i,i-1}$ $(i = 2, \cdots, l-h+1)$, which is in $\mathfrak{A}^{h+1}$ only if every $\alpha_{h-1+i,i} = 0$. That the first summation in (20) is zero now follows from $c_{h+1} = e_{h+1,h}d = \alpha_{hl} e_{h+1,1}$ which is in $\mathfrak{A}^{h+1}$ only if $\alpha_{hl} = 0$. Since $\mathfrak{A}^{h+1}$ contains

$$b_{l-h+2} = \sum_{k=l+1}^{n} \alpha_{k,l-h+2} e_{k,l-h+1}$$

it follows that every coefficient $\alpha_{k,l-h+2}$ is 0 and thus that $d = 0$.

FRESNO STATE COLLEGE AND PURDUE UNIVERSITY.

A THEORY OF SUBINVARIANT LIE ALGEBRAS.*

By EUGENE SCHENKMAN.**

A theory of subinvariant subgroups has been developed by Wielandt [7][1] who proved by way of application of the theory that the tower of automorphisms of a finite group with no center ends finitely. A similar study of subinvariant Lie algebras is presented here.

A subalgebra of a Lie algebra L is said to be subinvariant in L if it is a member of a sequence of subalgebras ending with L, each an ideal in the following. Some of the elementary properties of subinvariant subalgebras are given in Section 2. It is shown there that the intersection of two subinvariant subalgebras is subinvariant and that the property of subinvariance is preserved under homomorphisms; namely, if A is subinvariant in L, then $A\nu = \bar{A}$ is subinvariant in $\bar{L} = L\nu$, where ν is a homomorphism. Furthermore, if $\bar{A}$ is subinvariant in $\bar{L}$ and if A is the complete inverse image of $\bar{A}$ under ν, then A is subinvariant in L. We then show that if A and B are subinvariant in L, the composition factors from B to $A \cap B$ occur among those from L to A.

In Section 3 we prove that a subinvariant subalgebra is the sum of an ideal and a nilpotent algebra. This is done by proving that if A is subinvariant in L then the intersection $A^{\omega} = \bigcap_{k=1}^{\infty} A^k$ of all the power algebras of A is an ideal of L, and then by showing that an arbitrary algebra A is the sum of the ideal A^{ω} and a nilpotent algebra. These results are true for arbitrary base fields.

In Sections 4 through 7 we require that the base field be of characteristic zero. Section 4 is devoted to showing that the radical and nil radical of a subinvariant subalgebra are the intersections, respectively, of the subalgebra with the radical and nil radical of the containing algebra. The proof given is based on Cartan's decomposition theorem for a semisimple Lie algebra by means of which is is shown that the radical and nil radical are characteristic ideals.

In Section 5 it is shown that the algebra $\{A, B\}$, generated by two

---

* Received May 25, 1950. This paper is part of a dissertation presented for the degree of Doctor of Philosophy in Yale University.

** The author was an Atomic Energy Commission Predoctoral Fellow in the Physical Sciences of the National Research Council during the course of this research.

[1] Numbers in brackets refer to the bibliography.

subinvariant subalgebras A and B of L, is subinvariant in L; and that the different composition factors in a composition series from $\{A, B\}$ to A occur, not counting multiplicities of one dimensional factors, among those in a composition series from B to $A \cap B$. To show this last fact we prove that Lie algebra L can be expressed as the sum of its radical and the ideal $L^{(\omega)} = \bigcap_{k=1}^{\infty} L^{(k)}$, the intersection of all the derived algebras of L.

The problem of finding the smallest subinvariant subalgebra containing a given element of a Lie algebra L is considered in Section 6. It is shown that this subalgebra is the sum of an ideal and the one-dimensional algebra determined by the element itself.

In Section 7 we give several results relating an arbitrary subalgebra V of L to the greatest subinvariant subalgebra of L contained in V, and to the smallest subinvariant subalgebra of L containing V.

The last two section are concerned with an application of the theory to arbitrary Lie algebras. We show first in Section 8 that if A is a subinvariant subalgebra of L, and if $A^{\omega} = \bigcap_{k=1}^{\infty} A^{k}$ has a centralizer not contained in A^{ω}, then A has a non-zero centralizer in L. With this result we prove in Section 9 that the tower of derivation algebras, $L = D_0, D_1, \cdots, D_n, \cdots$ of a Lie algebra with no center over an arbitrary field ends finitely with $\dim D_n \leq d + c$, where c is the dimension of the center of $L^{\omega} = \bigcap_{k=1}^{\infty} L^{k}$, and d is the dimension of the derivation algebra of L^{ω}. As an immediate consequence of our tower theorem we have Chevalley's result [2] for characteristic zero that there exists an n such that D_n has only inner derivations.

It is of interest to contrast the theory of subinvariant Lie algebras with the theory of subinvariant groups. Almost without exception Wielandt's results [7] have analogues for Lie algebras; but there are important differences between the methods of the two theories. In the first place one has the concept of conjugate subgroups which is frequently used throughout the theory of subinvariant subgroups. This concept is available for nilpotent subinvariant subalgebras over fields of characteristic zero or for arbitrary subalgebras over the fields of real or complex numbers, but not in general. On the other hand the useful fact that if A is a subinvariant subalgebra of L, then $A^{\omega} = \bigcap_{k=1}^{\infty} A^{k}$ is an ideal of L, does not have an analogue for groups. Neither is there any analogue in group theory to Cartan's decomposition theorem for a semisimple Lie algebra which enables one to give the form of the minimal subinvariant subalgebra determined by a single element.

Before developing the theory I wish to express my sincere appreciation to Professor Nathan Jacobson for suggesting this research and for his constant interest and stimulating discussions during the course of its progress. I should also like to thank Professors Shizuo Kakutani and C. E. Rickart, and Mr. Charles Curtis for reading portions of the manuscript.

1. Preliminaries. We recall that L is a Lie algebra over a field Φ if L is a vector space over Φ in which the composition is defined such that for $l, m, n \varepsilon L$, $\alpha \varepsilon \Phi$

$$[l, l] = 0 \tag{1}$$

and

$$[[l, m], n] + [[m, n], l] + [[n, l], m] = 0 \tag{2}$$

in addition to the usual relations

$$[l, m + n] = [l, m] + [l, n]; \qquad [l + m, n] = [l, n] + [m, n]$$

and

$$\alpha[l, m] = [\alpha l, m] = [l, \alpha m].$$

We are concerned only with finite-dimensional algebras.

If A is a subspace of L and if $[a, l] \varepsilon A$ for every $a \varepsilon A$, $l \varepsilon L$, then A is an *ideal* or *invariant subalgebra* of L. It will be convenient to denote this by $A \lhd L$ or $L \rhd A$. A is a maximal ideal of L if and only if the difference alegbra $L - A$ is *simple.* A *normal series* of L is a series of algebras starting with (0) and ending with L, each an ideal in the following. If A and B are members of a normal series of L, A contained in B, then by a *normal series from A to B* we mean a series of algebras starting with A and ending with B, each an ideal of the following. Any normal series of L can be refined by addition of terms to be a *composition series* of L, that is, a normal series in which each algebra is a maximal ideal in the following. A *composition series from A to B* is similarly defined. The difference algebras of a normal series of L are called the *factors* of the normal series; those of a composition series of L are called *composition factors* or *simple factors* of L. From the Jordan-Hölder theorem [2] we know that the factors in a composition series from A to B are determined up to isomorphism.

If A and B are subspaces of L, then for $a \varepsilon A$, $b \varepsilon B$, $[A, B]$ denotes the spaces spanned by the products $[a, b]$. It is immediate from (1) and (2) that if A and B are ideals, then $[A, B]$ is an ideal. In particular the terms of the sequences

$$L \supseteq L^2 = [L, L] \supseteq L^3 = [L^2, L] \supseteq L^4 = [L^3, L] \supseteq \cdots$$

---

[2] See for instance Zassenhaus [8], p. 14.

and

$$L \supseteq L' = [L, L] \supseteq L'' = [L', L'] \supseteq L^{(3)} = [L'', L''] \supseteq \cdots$$

are ideals of L. The first of these is called the *power sequence* of L; the second, the *sequence of derived algebras* of L. If the power sequence ends with (0), the algebra is *nilpotent*. If the sequence of derived algebras ends with (0), the algebra is *solvable*. $L^{(n)} \subseteq L^{2^n}$ for $n = 1, 2, \cdots$ and hence a nilpotent algebra is always solvable. The composition factors of a Lie algebra are all one-dimensional if and only if the algebra is solvable. Subalgebras and difference algebras of solvable or nilpotent Lie algebras are solvable or nilpotent respectively.

It is well known that the join of two solvable ideals is a solvable ideal; and the join of two nilpotent ideals is a nilpotent ideal. Therefore every Lie algebra has a unique maximal solvable ideal called the *radical* and a unique maximal nilpotent ideal, the *nil radical*. An algebra with zero radical is said to be *semisimple*.

A Lie algebra L is *Abelian* if $L' = [L, L] = (0)$; hence an Abelian algebra is always solvable. The *center* of a Lie algebra L is the subalgebra consisting of all elements x such that $[x, L] = (0)$. The *centralizer* of a set of elements A of L is the set $Z(A)$ of elements x of L such that $[x, A] = (0)$. It is easy to check that $Z(A)$ is always an algebra and that if A is an ideal of L then so also is $Z(A)$. It follows that the center of a Lie algebra L is always an ideal. Any nilpotent Lie algebra has a non-zero center.

If L is a Lie algebra, a linear transformation $\tilde{h}$ acting on L is a *derivation* if for $a, b \in L$,

$$[a, b]\tilde{h} = [a\tilde{h}, b] + [a, b\tilde{h}]. \tag{3}$$

It is to be noted that the derivations of L form a Lie algebra if commutation $[\tilde{h}_1, \tilde{h}_2]$ is defined to be the linear transformation $\tilde{h}_1\tilde{h}_2 - \tilde{h}_2\tilde{h}_1$. For it is easy to verify that this mapping is a derivation. The algebra of all derivations of L will be denoted by $D(L)$.

If k is an element of L, then it follows from (1) and (2) that the *adjoint mapping* which maps a into $[a, k]$ for all $a \in L$ is a derivation; this will be called an *inner derivation* and will be denoted by ad k or $\tilde{k}$. By (1) and (2)

$$[a, [k, l]] = [[a, k], l] + [k, [a, l]] = a(\tilde{k}\tilde{l} - \tilde{l}\tilde{k})$$

for $a, k, l \in L$ and therefore the inner derivations form a subalgebra of $D(L)$. The mapping k into $\tilde{k}$ is a representation of L since it is a homomorphism of L into an algebra of linear transformations; it is called the *adjoint representation* of L. Obviously the kernel of the adjoint representation is the center of L. If L has no center then the adjoint representation is an isomorphism of L onto the algebra of inner derivations of L.

The mapping $\tilde{l}$ is *nilpotent in* L if $L\tilde{l}^m = (0)$ for some m. If $\tilde{l} = \operatorname{ad} l$ is nilpotent in L for every $l \varepsilon L$, then by Engel's theorem[3] L is nilpotent. If L is a solvable Lie algebra over a field of characteristic zero then Lie's theorem[4] states that $L' = [L, L]$ is nilpotent. Finally if L is a semisimple Lie algebra over a field of characteristic zero then by Cartan's theorem[5] $L = M_1 \oplus \cdot \cdot \cdot \oplus M_r$, where for $i = 1, \cdot \cdot \cdot, r$, M_i is a simple non-Abelian algebra which is an ideal of L.

2. Elementary properties of subinvariant subalgebras.

Definition. A is a *subinvariant subalgebra* of a Lie algebra L (or A is subinvariant in L) if there exist subalgebras A_i of L, $i = 0, 1, 2, \cdot \cdot \cdot, r$, such that $A = A_r$, $L = A_0$, and A_i is an ideal of A_{i-1}.

If A is subinvariant in L we write $A \lhd \lhd L$ or $L \rhd \rhd A$. It is obvious that a subalgebra is subinvariant if and only if it is a member of a normal series of L.

Some easy consequences of the definition of subinvariance are given in

THEOREM 1. *Let A and B be subinvariant in L, C subinvariant in A and let L^* be any subalgebra of L. Then*

(a) *C is subinvariant in L.*

(b) *$A \cap L^*$ is subinvariant in L^*.*

(c) *If $A \subseteq L^*$, then A is subinvariant in L^*.*

(d) *$A \cap B$ is subinvariant in L.*

(e) *If ν is a homomorphism of L onto $\bar{L}$, then $A\nu = \bar{A}$ is subinvariant in $\bar{L}$.*

(f) *Conversely, if $\bar{F}$ is subinvariant in $\bar{L}$, and if F is the complete inverse image of $\bar{F}$ under ν, then F is subinvariant in L.*

(g) *If L is nilpotent, then L^* is subinvariant in L.*

Proof. (a) follows directly from the definition of subinvariant subalgebra; for $C = C_s \lhd C_{s-1} \lhd \cdot \cdot \cdot \lhd C_1 \lhd A = A_r \lhd \cdot \cdot \cdot \lhd A_1 \lhd L$. To prove

---

[3] See Jacobson [4], p. 881.

[4] *Ibid.*, p. 880.

[5] See Cartan [1], p. 53; or Weyl [6], pp. 363 ff.

(b) we have $L = A_0 \triangleright A_1 \triangleright \cdots \triangleright A_r = A$ and we let $(L^* \cap A_i) = A_i^*$ for $i = 0, 1, 2, \cdots, r$. Then since $A_j \triangleright A_{j+1}$, it follows that $A_j^* \triangleright A_{j+1}^*$, $j = 0, 1, \cdots, r-1$. For if $a_j^* \varepsilon A_j^*$, $a_{j+1}^* \varepsilon A_{j+1}^*$, then $[a_j^*, a_{j+1}^*] \varepsilon A_{j+1}^* = (L^* \cap A_{j+1})$. Therefore $L^* = A_0^* \triangleright A_1^* \triangleright \cdots \triangleright A_r^* = (A \cap L^*)$, proving (b). Also, (c) follows immediately from (b); and (d) follows from (c) with (a).

To prove (e) we have $L = A_0 \triangleright A_1 \triangleright \cdots \triangleright A_r = A$. If for $i = 0, 1, 2, \cdots, r$, $A_i\nu$ is denoted by $\bar{A}_i$, then $\bar{A}_i \triangleright \bar{A}_{i+1}$. For if $\bar{a}_j \varepsilon \bar{A}_j$, where $j = i$, $i + 1$, then $\bar{a}_j = a_j\nu$ for some $a_j \varepsilon A_j$ and $[\bar{a}_i, \bar{a}_{i+1}] = [a_i\nu, a_{i+1}\nu] \varepsilon \bar{A}_i$ since $[a_i, a_{i+1}] \varepsilon A_i$. It follows that $\bar{L} = \bar{A}_0 \triangleright \bar{A}_1 \triangleright \cdots \triangleright \bar{A}_r = \bar{A}$. Conversely, to prove (f), if $\bar{L} = \bar{F}_0 \triangleright \bar{F}_1 \triangleright \cdots \triangleright \bar{F}_r = \bar{F}$, let F_i be the complete inverse image of $\bar{F}_i$ for $i = 0, 1, \cdots, r$, and let $f_i \varepsilon F_i$. Then since $[f_i, f_{i+1}]\nu = [f_i\nu, f_{i+1}\nu] \varepsilon \bar{F}_{i+1}$, it follows that $[f_i, f_{i+1}] \varepsilon F_{i+1}$ and consequently $F_{i+1} \triangleleft F_i$. Then $L = F_0 \triangleright F_1 \triangleright \cdots \triangleright F_r = F$ and $F \triangleleft \triangleleft L$, as was to be proved.

Finally, to prove (g), let L be nilpotent. Then $L^k = 0$ for some k. Consequently $L^* \triangleleft L^* + L^{k-1} \triangleleft L^* + L^{k-2} \triangleleft \cdots \triangleleft L^* + L = L$, and $L^* \triangleleft \triangleleft L$. This completes the proof of all parts of the theorem.

LEMMA 1. *If M is a maximal ideal of L and if A is subinvariant in L, then either $A \subseteq M$ or $A \cap M$ is a maximal ideal of A and $A - (A \cap M) \cong L - M$. (Here $\cong$ stands for isomorphic, $\sim$ for homomorphic onto.)*

Proof. Let $F = L - M$. Then $L \sim F$, where F is simple. Since composition series map into composition series under homomorphism, either A maps into (0), in which case $A \subseteq M$, or A maps onto F and then $(A + M) - \mathrm{M} \cong F = L - M$. It follows from the Second Isomorphism Theorem[6] that $A - (A \cap M) \cong L - M$, proving the lemma.

As a consequence we have

THEOREM 2. *If A and B are subinvariant subalgebras of L, then the factors in a composition series from B to $A \cap B$ appear among those in a composition series from L to A.*

Proof. $L = A_0 \triangleright A_1 \triangleright \cdots \triangleright A_r = A$, where for $i = 1, 2, \cdots, r$, A_i is a maximal ideal of A_{i-1}. Hence $B = (B \cap A_0) \trianglerighteq (B \cap A_1) \trianglerighteq \cdots \trianglerighteq (B \cap A_r) = (B \cap A)$ and the theorem is obvious.

3. A subinvariant subalgebra as the sum of an ideal and a nilpotent algebra.

A key result that will be used throughout our theory of subinvariant Lie algebras is stated in

---

[6] See Zassenhaus [8], p. 12.

THEOREM 3. *If A is a subinvariant subalgebra of a Lie algebra L, then $A^\omega = \bigcap_{i=1}^{\infty} A^i$ is an ideal of L.*

Proof. $A^\omega = A^p$ for some integer p because of the finite dimensionality of L. It must be shown that $A^p \lhd L$. We know that $[L, A^2] \subseteq [[L, A], A]$ by (1) and (2), and we proceed by induction. If we assume for every subspace B of L that

$$[B, A^k] \subseteq [\cdots[[B, \overbrace{A], A], \cdots, A}^{k \text{ times}}]$$

then it follows that

$$\begin{aligned}
[B, A^{k+1}] &= [B, [A, A^k]] \subseteq [[B, A], A^k] + [[B, A^k], A] \\
&\subseteq [\cdots[[B, A], \overbrace{A], \cdots, A}^{k \text{ factors}}] + [[\cdots[B, \overbrace{A], \cdots, A}^{k \text{ factors}}]A] \\
&= [\cdots[B, \overbrace{A], \cdots, A}^{k+1 \text{ factors}}].
\end{aligned}$$

Hence for all n,

$$[L, A^n] \subseteq [\cdots[L, \overbrace{A], \cdots, A}^{n \text{ factors}}].$$

If n is chosen equal to $p + q$, where q is the number of factors in a composition series from L to A, then the theorem is proved.

COROLLARY 1. *If A is a subinvariant subalgebra of L such that $A = A'$, then A is an ideal of L.*

Remark. The following argument shows that the analogous statement is not true for finite groups.

Let A and B be isomorphic finite groups and consider the direct product G of A and B. Then the elements of G can be written uniquely in the form $a_i b_j$ where $a_i \in A$, $b_j \in B$. If a_j and b_j are corresponding elements under the isomorphism between A and B then the mapping $a_i b_j \rightarrow a_j b_i$ determines an automorphism of G which we will denote by σ. Obviously σ is of order 2 and if we form the subgroup H of the holomorph generated by σ and G, then G is of index 2 in H and hence is normal in H. Then it is clear that σ transforms A into B. That is, A is not normal in H. On the other hand A is normal in G and G is normal in H. Hence A is subinvariant but not invariant in H. This construction is available for an arbitrary group, in particular if the group $A = [A, A]$, where $[A, A]$ is the commutator subgroup of A, and hence we see that the analogue to Corollary 1 is not valid for finite groups.

We return to the theory of Lie algebras.

Our next task is to obtain an expression for an arbitrary subinvariant subalgebra as the sum of an ideal and a nilpotent subalgebra. To do so we first review briefly some general concepts.

If L is a vector space and $\tilde{h}$ an arbitrary linear transformation on L, we put $C = \bigcap_{i=1}^{\infty} L\tilde{h}^i$ and let $B = \bigcup_{i=1}^{\infty} B_i$, where $B_i = \{x \mid x\tilde{h}^i = 0\}$ is the linear subspace of L consisting of all elements x such that $x\tilde{h}^i = 0$. Then $C = C\tilde{h}$, $B\tilde{h}^n = 0$ for some n because of the finite dimensionality, and $L = B + C$.[7] The spaces C and B will be called the *preserved space* and *nil* space, respectively, of L relative to $\tilde{h}$, or briefly of $\tilde{h}$ if there is no danger of ambiguity. If in particular L is a Lie algebra and $\tilde{h}$ a derivation, then the nil space B of $\tilde{h}$ is an algebra. For if $b_i \varepsilon B$ for $i = 1, 2$, then $b_i\tilde{h}^{s_i} = 0$ for some s_i. It follows from Leibniz's rule that

$$[b_1,b_2]\tilde{h}^{s_1+s_2-1} = \sum_{t=0}^{s_1+s_2-1} \binom{s_1+s_2-1}{t} [b_1\tilde{h}^t, b_2h^{s_1+s_2-1-t}] = 0$$

since either $t \geq s_1$ or $s_1 + s_2 - 1 - t \geq s_2$. Hence $[b_1, b_2] \varepsilon B$ and B is an algebra as asserted. B is called the *nil algebra* of L relative to $\tilde{h}$, or briefly, of $\tilde{h}$.

It is now possible to prove

THEOREM 4. *If L is a Lie algebra, then $L = L^{\omega} + H$, where H is a nilpotent algebra and $L^{\omega} = \bigcap_{i=1}^{\infty} L^i$.*

Proof. Let n be the dimension of L and assume the theorem true for algebras of dimension less than n. If $L\tilde{l}^n = 0$ for all $l \varepsilon L$, where $\tilde{l} = \operatorname{ad} l$, then L is nilpotent by Engel's theorem, and there is nothing to prove.

Otherwise there is some $l_0 \varepsilon L$ such that l_0 is not nilpotent in L. By the above remarks $L = B + C$, where C is the nil algebra of l_0 with dimension less than n and B is the preserved space. Since $B\tilde{l}_0 = B$, it follows that $B \subseteq L^{\omega}$. On the other hand by the induction assumption $C = C^{\omega} + H$, where H is a nilpotent algebra. But $C^{\omega} \subseteq L^{\omega}$. Therefore $L = B + C = B + C^{\omega} + H = L^{\omega} + H$ because $B + C^{\omega} \subseteq L^{\omega}$, and the theorem is proved.

Combining the two previous theorems we have the desired result.

THEOREM 5. *If A is a subinvariant subalgebra of a Lie algebra L, then $A = A^{\omega} + H$, where $A^{\omega} = \bigcap_{i=1}^{\infty} A^i$ is an ideal of L, and H is a nilpotent algebra.*

---

[7] See Jacobson [5], p. 9.

We could now pass immediately to Sections 8 and 9 and prove the tower theorem. Before doing so, however, we will first complete the general theory.

4. The radical and nil radical of a subinvariant subalgebra. This and the following three sections will deal with Lie algebras over a field of characteristic zero.

LEMMA 2. *A solvable subinvariant subalgebra of a Lie algebra over a field of characteristic zero is contained in the radical of the algebra.*

Proof. Let L be the Lie algebra with radical R and solvable subinvariant subalgebra S. If $\bar{L} = L - R$, then $\bar{L}$ is semisimple and by Cartan's theorem $\bar{L} = \bar{M}_1 \oplus \cdots \oplus \bar{M}_n$, where for $i = 1, \cdots, n$, $\bar{M}_i$ is a simple non-Abelian ideal of $\bar{L}$. It follows from the form of $\bar{L}$ that as a subinvariant subalgebra of $\bar{L}$, $\bar{S} = (S + R) - R = \bar{M}_i \oplus \cdots \oplus \bar{M}_{i_p}$, $1 \leq i_1 \leq \cdots < i_p < n$. But $\bar{S}$ is solvable and has no non-Abelian factors. Therefore $\bar{S} = 0$ and $S \subseteq R$ as the lemma asserts.

The next lemma gives results which are well known,[8] but the proof is inserted here for completeness.

LEMMA 3. *The radical R and nil radical N of a Lie algebra L over a field of characteristic zero are left invariant by every derivation $\tilde{d}$ of L; and in fact $R\tilde{d} \subseteq N$.*

Proof. Let $\tilde{d}$ be a derivation of L and form the Lie algebra $\tilde{L} = L + \tilde{d}$ where multiplication between $\tilde{d}$ and $l \in L$ is defined to be $[l, \tilde{d}] = l\tilde{d}$. From this it is obvious that $L \lhd \tilde{L}$. Let $\tilde{R}$ be the radical of $\tilde{L}$. Then $R \subseteq \tilde{R} \cap L$; for $R \subseteq L$, and by the preceding lemma $R \subseteq \tilde{R}$ since $R \lhd L \lhd \tilde{L}$. On the other hand $\tilde{R} \cap L$ is an ideal of $\tilde{L}$ since both $\tilde{R}$ and L are ideals of $\tilde{L}$. It follows that $\tilde{R} \cap L$ is an ideal of L. Since $\tilde{R}$ is solvable, $\tilde{R} \cap L$ is a solvable ideal of L and $\tilde{R} \cap L \subseteq R$. Hence $R = \tilde{R} \cap L$. As we have seen, this implies that R is an ideal of $\tilde{L}$ and therefore $R\tilde{d} \subseteq R$; and thus the radical is invariant under every derivation $\tilde{d}$ of L, as the first part of the lemma asserts.

To show that the nil radical N is invariant with respect to the derivation $\tilde{d}$ we consider the subalgebra $R_1 = R + \tilde{d}$ of $L + \tilde{d} = \tilde{L}$. This is a subalgebra since it has already been shown that $R\tilde{d} \subseteq R$. It follows that R_1 is solvable and consequently by Lie's theorem $R_1' = [R_1, R_1]$ is nilpotent.

---

[8] The proof given here was presented by Professor N. Jacobson to his class in non-associative algebras at Yale University. See also Hochschild [3], p. 62.

But $R_1' \lhd R_1$; hence $R_1' \lhd R$ since $R_1' \subseteq R$. Then as a nilpotent ideal of R, $R_1' \subseteq N_R$, the nil radical of R. Hence $R\tilde{d} \subseteq R_1' \subseteq N_R$. By taking $\tilde{d}$ to be an inner derivation determined by the element l of L we have $[R, l] \subseteq N_R$. Since this is true for all $l \varepsilon L$, $[R, L] \subseteq N_R$. Hence $[N_R, L] \subseteq N_R$ and N_R is a nilpotent ideal of L. Then by the definition of nil radical $N_R \subseteq N$. Obviously $N_R \supseteq N$ and thus $N = N_R$. But we had $R\tilde{d} \subseteq N_R$. Hence $R\tilde{d} \subseteq N$ and consequently also $N\tilde{d} \subseteq N$, completing the proof of the lemma.

LEMMA 4. *If M is a nilpotent subinvariant subalgebra of L, a Lie algebra over a field of characteristic zero, then M is contained in N, the nil radical of L.*

Proof. $M \lhd \lhd L$ and therefore $M \lhd M_1 \lhd M_2 \lhd \cdots \lhd M_n = L$. For $j = 1, \cdots, n$, let N_j be the nil radical of M_j and let $m_j \varepsilon M_j$. Then $M \subseteq N_1$ since M is a nilpotent ideal of M_1. But for $i = 1, \cdots, n-1$, $[M_i, m_{i+1}] \subseteq M_i$ and therefore m_{i+1} defines a derivation $\tilde{m}_{i+1}$ in M_i. It follows from the preceding lemma that $N_i\tilde{m}_{i+1} \subseteq N_i$, whence $N_i \lhd M_{i+1}$ and consequently $N_i \subseteq N_{i+1}$. Then $M \subseteq N_1 \subseteq N_2 \subseteq \cdots \subseteq N_n = N$, proving the lemma.

We can now prove the main result of this section.

THEOREM 6. *Let L be a Lie algebra over a field of characteristic zero, with radical R and nil radical N. Let A be a subinvariant subalgebra of L with radical R_A and nil radical N_A. Then $R_A = A \cap R$ and $N_A = A \cap N$.*

Proof. $A \cap R$ is a solvable ideal of A since $R \lhd L$ and hence $A \cap R \subseteq R_A$. On the other hand $R \subseteq A$ and, by Lemma 2, $R_A \subseteq R$. Hence $R_A \subseteq A \cap R$ and $R_A = A \cap R$. In like manner Lemma 4 gives that $N_A = A \cap N$ and the proof is complete.

5. The sum theorems.

THEOREM 7. *Let A and B be subinvariant subalgebras of L, a Lie algebra over a field of characteristic zero. If $\{A, B\}$ denotes the algebra generated by A and B, then $\{A, B\}$ is subinvariant in L.*

Proof. If both A and B are nilpotent then, by Lemma 4, A, B and consequently $\{A, B\}$ are contained in the nil radical N of L. Then by Theorem 1 (g), $\{A, B\}$ is subinvariant in N and therefore in L. Thus in this case the theorem is true.

In a general case we have $(0) \neq A^\omega$ or $(0) \neq B^\omega$. Therefore, by

Theorem 3, $\{A^\omega, B^\omega\}$ is a non zero ideal of L. Let $\bar{A} = (A + \{A^\omega + B^\omega\}) - \{A^\omega, B^\omega\}$, $\bar{B} = (B + \{A^\omega, B^\omega\}) - \{A^\omega, B^\omega\}$. Then $\bar{A}$ and $\bar{B}$ are nilpotent in $\bar{L} = L - \{A^\omega, B^\omega\}$ and by the first part of the proof $\{\bar{A}, \bar{B}\} \lhd \lhd \bar{L}$. But $\{\bar{A}, \bar{B}\} = \overline{\{A, B\}}$ where $\overline{\{A, B\}} = \{A, B\} - \{A^\omega, B^\omega\}$. For if C_i stands for A or B, then $\overline{\{A, B\}} = \overline{\sum[\cdots[C_1, C_2], \cdots, C_n]} = \sum[\cdots[\bar{C}_1, \bar{C}_2], \cdots, \bar{C}_n] = \{\bar{A}, \bar{B}\}$. Hence $\{\bar{A}, \bar{B}\} = \overline{\{A, B\}}$ and by the above $\overline{\{A, B\}} \lhd \lhd \bar{L}$. It follows from Theorem 1 (f) that $\{A, B\} \lhd \lhd L$ and the proof is complete.

In like manner we prove the following result.

Theorem 8. *Let A be subinvariant in L, a Lie algebra over a field of characteristic zero. Let b be any element of L and let A^* be the smallest algebra containing A such that $[A^*, b] \subseteq A^*$. Then A^* is subinvariant in L.*

Proof. We first note that $A^* = \{\sum_{k=0}^{\infty} A\tilde{b}^k\}$. For certainly $\{\sum_{k=0}^{\infty} A\tilde{b}^k\}$ contains A and is invariant relative to $\tilde{b}$ since the space spanned by $\sum_{k=0}^{\infty} A\tilde{b}^k$ is left invariant by $\tilde{b}$. Hence $\{\sum_{k=}^{\infty} A\tilde{b}^k\} \supseteq A^*$. On the other hand $A^* \supseteq A$ and since $[A^*, b] \subseteq A^*$, $A^* \supseteq A\tilde{b}^k$ for all k. It follows that $A^* \supseteq \{\sum_{k=0}^{\infty} A\tilde{b}^k\}$. Consequently $A^* = \{\sum_{k=0}^{\infty} A\tilde{b}^k\}$.

If now A is nilpotent then $A \subseteq N$, the nil radical of L. But N is an ideal of L and therefore by its form $A^* \subseteq N$ and consequently A^* is subinvariant in L. If A is not nilpotent then A^ω is a non zero ideal of L. Let $\bar{L} = L - A^\omega$, $\bar{A} = A - A^\omega$, $\overline{A^*} = A^* - A^\omega$, and $\bar{b} = b + A^\omega$. Then $\overline{A^*} = \{\sum_{k=0}^{\infty} A\tilde{b}^k\} = \{\sum_{k=0}^{\infty} \bar{A}\tilde{\tilde{b}}^k\}$. But $\bar{A}$ is in the nil radical of $\bar{L}$ and hence so also is $\{\sum_{k=0}^{\infty} \bar{A}\tilde{\tilde{b}}^k\} = A^*$. It follows that $\overline{A^*} \lhd \lhd \bar{L}$ and hence $A^* \lhd \lhd L$, proving the theorem.

From Theorem 2 we know that if A and B are subinvariant then the factors in a composition series between A and $(A \cap B)$ occur among those in a composition series between $\{A, B\}$ and B. The converse of this, namely that the factors in a composition series from $\{A, B\}$ to B occur among those in a composition series from A to $(A \cap B)$ is true if we add the qualifying statement "not counting multiplicities of one-dimensional factors," and if we assume that the base field has characteristic zero. To show this we need the following result.

Lemma 5. *If L is a Lie algebra over a field of characteristic zero, then $L = L^{(\omega)} + R$, where $L^{(\omega)} = \bigcap_{i=1}^{\infty} L^{(i)}$ and R is the radical of L.*

Proof. $L - R$ is semisimple. Therefore $L - R = [L - R, L - R]$

$= \cdot \cdot \cdot = (L - R)^{(\omega)}$ and $L - R = (L^{(\omega)} + R) - R$. Consequently $L = L^{(\omega)} + R$ as was to be shown.

It is interesting to contrast this lemma with Theorem 5. If A is subinvariant in L, and if R_A is the radical of A, then the lemma tells us that A is the sum of two algebras which are subinvariant in L. On the other hand if the properties of A alone are being considered, then the expression $A^{\omega} + H$ given by Theorem 5 is useful because of the nilpotency of H.

We will now prove the partial converse of Theorem 2.

THEOREM 9. *If A and B are subinvariant in L, a Lie algebra over a field of characteristic zero, then the factors in a composition series from $\{A, B\}$ to A occur among those in a composition series from B to $(A \cap B)$, not counting multiplicities of one-dimensional factors.*

Proof. We observe first that the composition factors are never one-dimensional or Abelian between a Lie algebra and its radical and are always one-dimensional for a solvable algebra. Now by the corollary to Theorem 3, $A^{(\omega)} = \bigcap_{i=1}^{\infty} A^{(i)}$ is an ideal of L. Let $\bar{L} = L - A^{(\omega)}$, $\bar{A} = A - A^{(\omega)}$, $\bar{B} = (B + A^{(\omega)}) - A^{(\omega)}$, $\overline{\{A, B\}} = \{A, B\} - A^{(\omega)}$, and $\overline{(A \cap B)} = ((A \cap B) + A^{(\omega)}) - A^{(\omega)}$. We note that, as in the proof of Theorem 7, $\overline{\{A, B\}} = \{\bar{A}, \bar{B}\}$ and $\overline{(A \cap B)} = \bar{A} \cap \bar{B}$. Then it is well known that the composition factors from $\{A, B\}$ to A are the same as those from $\overline{\{A, B\}}$ to $\bar{A}$. Also the factors between $\bar{B}$ and $\overline{(A \cap B)}$ occur among those between B and $A \cap B$. Hence it is sufficient to prove the theorem for $\bar{A}$, $\bar{B}$ in $\bar{L}$.

In this case $\bar{A}$ and consequently $(\bar{A} \cap \bar{B})$ are solvable. By the preceding lemma $\bar{B} = \bar{B}^{(\omega)} + \bar{R}_B$ where $\bar{R}_B$ is the radical of $\bar{B}$ and $\bar{B}^{(\omega)} = \bigcap_{i=1}^{\infty} \bar{B}^{(i)}$. Then $\bar{R}_B$ and $\bar{A}$ are contained in $\bar{R}$, the radical of $\bar{L}$. Hence $\overline{\{A, B\}} = \{\bar{A}, \bar{B}\} = \{\bar{A}, \bar{B}^{(\omega)}, \bar{R}_B\} \subseteq \{\bar{B}^{(\omega)}, \bar{R}\} = \bar{B}^{(\omega)} + \bar{R}$ since each of the latter algebras is an ideal of $\bar{L}$. Now the non-Abelian composition factors of $\bar{B}^{(\omega)} + \bar{R}$ are the same as those of $(\bar{B}^{(\omega)} + \bar{R}) - \bar{R}$, and consequently are contained in $\bar{B}^{(\omega)}$ since $(\bar{B}^{(\omega)} + \bar{R}) - \bar{R}$ is a homomorphic image of $B^{(\omega)}$. It follows that the non-Abelian composition factors of $\{\bar{A}, \bar{B}\}$ must occur among those of $\bar{B}^{(\omega)}$, consequently in $\bar{B}$; and hence between $\bar{B}$ and $\overline{(A \cap B)}$, since $\overline{(A \cap B)}$ is solvable, and therefore in $\bar{R}_B$.

There are two possible cases in considering the one-dimensional factors. First there may be one or more one-dimensional factors between $\bar{B}$ and $(\bar{A} \cap \bar{B})$; in this case the theorem is obviously true. On the other hand, if none of the factors from $\bar{B}$ to $(\bar{A} \cap \bar{B})$ is one-dimensional then $\bar{R}_B = (\bar{A} \cap \bar{B})$, since the latter is solvable and subinvariant and hence contained in $\bar{R}_B$. In this case, $\{\bar{A}, \bar{B}\} = \{\bar{A}, \bar{B}^{(\omega)}, \bar{R}_B\} = \{\bar{A}, \bar{B}^{(\omega)}\} = \bar{A} + \bar{B}^{(\omega)}$, since $\bar{B}^{(\omega)}$ is an ideal of $\bar{L}$. Then $\bar{A}$ is the radical of $\bar{A} + \bar{B}^{(\omega)}$. For certainly $\bar{A}$ is contained in $\bar{S}$, the radical of $\bar{A} + \bar{B}^{(\omega)}$. On the other hand, if $\bar{s} \varepsilon \bar{S}$, then $\bar{s} = \bar{a} + \bar{b}$, where $\bar{a} \varepsilon \bar{A}$, $\bar{b} \varepsilon \bar{B}^{(\omega)}$. But then since $\bar{a} \varepsilon \bar{S}$, $\bar{b} \varepsilon \bar{S}$ and $\bar{b} \varepsilon \bar{R}_B \subseteq \bar{A}$; hence $\bar{S} \subseteq \bar{A}$. It follows that $\bar{A}$ is the radical of $\bar{A} + \bar{B}^{(\omega)} = \{\bar{A}, \bar{B}\}$ and the factors from $\{\bar{A}, \bar{B}\}$ to $\bar{A}$ are non-Abelian. This completes the proof of the theorem.

6. The minimal subinvariant subalgebra determined by a single element.

It is clear that there is a unique minimal subinvariant subalgebra determined by an arbitrary element of a Lie algebra because of the finite dimensionality and because the intersection of two subinvariant algebras is subinvariant.

We shall consider first the minimal subinvariant subalgebra determined by an element of the radical. For this we need the following result.

LEMMA 6. *Let a be an element of the Lie algebra L over an arbitrary field, and let L_1 be the preserved space of $\tilde{a}$. If $a \varepsilon A$, A subinvariant in L, then L_1 is contained in A.*

Proof. $L_1 = L_1\tilde{a}^k$, which for k large enough is in A, since $a \varepsilon A$, and A is subinvariant in L.

We can now prove the following theorem.

THEOREM 10. *Let R be the radical of L, a Lie algebra over a field of characteristic zero, and let L_1 be the preserved space in L of the derivation $\tilde{a}$ determined by an element a in R. Let L_a be the algebra generated by L_1 and let $L_a + (a) = A$. Then A is the minimal subinvariant subalgebra of L containing a, and L_a is an ideal of L. L_a will be called the* associated ideal *to a.*

Proof. In order to prove that A is an algebra it is sufficient to observe that if $\tilde{a} = \operatorname{ad} a$, then $L_1\tilde{a} = L_1$ and consequently $L_a\tilde{a} \subseteq L_a$. We observe next that $L_1 = L_1\tilde{a} \subseteq R$, since $a \varepsilon R$ and R is an ideal. It follows that $L_1 = L_1\tilde{a} \subseteq R' \subseteq N$, the nil radical of L, and consequently $L_a \subseteq N$. Then for all s, $[N^s, L_a] \subseteq N^{s+1}$. Let $N^s = N_{0,s} \oplus N_{1,s}$ where $N_{0,s}$ is the nil space, $N_{1,s}$ the preserved space of N^s relative to $\tilde{a}$. Then $N_{1,s} \subseteq L_1 \subseteq L_a$ and $N_{0,s}\tilde{a}^{k_s} = 0$ for k_s depending on s. If $N^t = 0$ we have

$$\begin{aligned} L_a + (a) = A &\lhd A + N_{0,t-1}\tilde{a}^{k_{t-1}-1} \lhd A + N_{0,t-1}\tilde{a}^{k_{t-1}-2} \\ &\lhd \cdots \lhd A + N_{0,t-1} = A + N^{t-1} \lhd A + N^{t-1} + N_{0,t-2}\tilde{a}^{k_{t-2}-1} \\ &\lhd \cdots \lhd A + N \lhd R \lhd L \end{aligned}$$

where each step is true because of the facts that $A = L_a + (a)$, $[N^s, L_a] \subseteq N^{s+1}$, and $[N_{0,t-r}\tilde{a}^{k_{t-r}-p_r}, a] = N_{0,t-r}\tilde{a}^{k_{t-r}-p_r+1}$ for $0 < r < t$ and $0 < p_r < k_{t-r}$. This proves that A is subinvariant in L. A is the minimal subinvariant subalgebra containing a by the previous lemma.

To show that L_a is an ideal of L we show that $L_a = \bigcap_{i=1}^{\infty} A^i$, as follows. $L_a \supseteq \bigcap_{i=1}^{\infty} A^i$; for $L_a \supseteq A^2$, since $A = L_a + (a)$ and $L_a\tilde{a} \subseteq L_a$. On the other hand, $L_1 = L_1\tilde{a} \subseteq \bigcap_{i=1}^{\infty} A^i$ and consequently $L_a \subseteq \bigcap_{i=1}^{\infty} A^i$, since L_a is the algebra generated by L_1. Then by Theorem 3, L_a is an ideal of L, and the proof of the theorem is complete.

We observe that if $a \in N$, then $L_1 = 0$, $A = (a)$ and the theorem is true by virtue of Theorem 1 (g).

The next task is to study the subinvariant subalgebras which map onto the simple components of a Lie algebra modulo its radical. The result that we shall need is contained in the following theorem.

THEOREM 11. *Let L be a Lie algebra over a field of characteristic zero with radical R; and let $\bar{L} = L - R = \sum_{i=1}^{n} \oplus \bar{M}_i$, where for $i = 1, \cdots, n$, $\bar{M}_i$ is a simple algebra and an ideal of $\bar{L}$. If ν is the natural mapping of L onto $\bar{L}$, let T_i be the complete inverse image under ν of $\bar{M}_i$. Then $T_i^{(\omega)} = \bigcap_{k=1}^{\infty} T_i^{(k)}$ is an ideal of L such that $T_i^{(\omega)}\nu = \bar{M}_i$, and $T_i^{(\omega)}$ is contained in every subvariant subalgebra of L which maps under ν onto $\bar{M}_i$.*

Proof. Since $\bar{M}_i$ is an ideal of $\bar{L}$, it follows that T_i and consequently $T_i^{(\omega)}$ are ideals of L. Now, T_i contains R and R is the radical of T_i. By its definition $T_i - R = \bar{M}_i$, and therefore in a composition series of T_i there is only one simple non-Abelian factor, and this factor is isomorphic to $\bar{M}_i$. It follows that if A_i is any subinvariant subalgebra of T_i, then a simple non-Abelian factor must occur either in a composition series from T_i to A_i or in that of A_i, but not in both series. In the former case A_i is solvable and hence is contained in R; in the latter case A_i is not solvable and $A_i\nu = \bar{M}_i$.

Suppose now that A_i and B_i are subinvariant subalgebras of L which map under ν onto $\bar{M}_i$. Then A_i and B_i are contained in T_i, since T_i is the complete inverse image of $\bar{M}_i$; and the simple factors between T_i and A_i and between T_i and B_i are all one-dimensional, since $A_i\nu = B_i\nu = \bar{M}_i$. By Theorem 2 it follows that the simple factors between A_i and $(A_i \cap B_i)$ are one-dimensional and hence the same result holds for the simple factors between T_i and $(A_i \cap B_i)$. It follows that $(A_i \cap B_i)\nu = \bar{M}_i$. From this together with the facts that $(A_i \cap B_i)$ is subinvariant in L and that L is finite-dimensional we know that there exists a subinvariant subalgebra V_i of L such that $V_i\nu = \bar{M}_i$ and such that $V_i \subseteq A_i$ if A_i is any subinvariant subalgebra of L whose image under ν is $\bar{M}_i$. Now, $T_i^{(\omega)}\nu = \bar{M}_i$ since the simple factors between T_i and $T_i^{(\omega)}$ are one-dimensional and $T_i^{(\omega)}$ is subinvariant, in fact, an ideal of L. Hence $T_i^{(\omega)} \supseteq V_i$. But then the factors between $T_i^{(\omega)}$ and V_i are one-dimensional since $T_i^{(\omega)}\nu = V_i\nu = \bar{M}_i$. This implies that $T_i^{(\omega)} = V_i$ and the theorem is proved.

We note that $(\sum_{i=1}^{n} T_i^{(\omega)})\nu = \sum_{i=1}^{n} \bar{M}_i = \bar{L}$ and hence the composition factors between L and $\sum_{i=1}^{n} T_i^{(\omega)}$ are all one-dimensional. It follows that $L^{(\omega)} \subseteq \sum_{i=1} T_i^{(\omega)}$. On the other hand $(\sum_{i=1} T_i^{(\omega)})' = \sum_{i=1} T_i^{(\omega)}$ and hence $L^{(\omega)} = \sum_{i=1}^{n} T_i^{(\omega)}$. Then from Lemma 5 we also have that $L = \sum_{i=1}^{n} T_i^{(\omega)} + R$.

If now l is arbitrary in L, let $l\nu = \sum_{j \varepsilon J} \bar{m}_j$, where $J = (i_1, \cdots, i_r)$ and where $\bar{m}_j \,\varepsilon\, \bar{M}_j$, $\bar{m}_j \neq 0$. It is possible to choose $b \,\varepsilon \sum_{j \varepsilon J} T_j^{(\omega)}$ such that $b\nu = l\nu$, since $\sum_{j \varepsilon J} T_j^{(\omega)}$ maps onto $\sum_{j \varepsilon J} \bar{M}_j$. Hence $l - b = a \,\varepsilon\, R$ and we let L_a be the associated ideal of a. The characterization of the minimal subinvariant subalgebra determined by l is then given in the following theorem.

THEOREM 12. *Let L be a Lie algebra satisfying the hypotheses of the preceding theorem. Let A be the minimal subinvariant subalgebra determined by an arbitrary element l of L. Then in the notation above,*

$$A = \sum_{j \varepsilon J} T_j^{(\omega)} + L_a + (a).$$

Proof. $A\nu$ contains $l\nu = \sum_{j \varepsilon J} \bar{m}_j$, where $\bar{m}_j \neq 0$. Hence $A\nu$ contains $\sum_{j \varepsilon J} \bar{M}_j$, since the only ideals and therefore the only subinvariant subalgebras of a semisimple algebra are sums of the simple components. It follows from Theorem 11 that A contains $\sum_{j \varepsilon J} T_j^{(\omega)}$. Then $b \,\varepsilon\, A$ and consequently $l - b = a \,\varepsilon\, A$. Then by Theorem 10, $L_a \subseteq A$. Therefore $A \supseteq \sum_{j \varepsilon J} T_j^{(\omega)} + L_a + (a)$. But the latter expression represents a subinvariant algebra containing l. For $L_a + (a)$ is subinvariant by Theorem 10; and $\sum_{j \varepsilon J} T_j^{(\omega)}$ is an ideal. Hence by the minimality of A, $A = \sum_{j \varepsilon J} T_j^{(\omega)} + L_a + (a)$, and the theorem is proved.

7. The greatest subinvariant subalgebra contained in, and the least subinvariant subalgebra containing, an arbitrary algebra. If U is an arbitrary subalgebra of a Lie algebra L, let U_* denote the algebra generated by all the subinvariant subalgebras of L contained in U; and let U^* denote the intersection of all the subinvariant subalgebras of L containing U. Then by Theorem 1 (d), U^* is minimal such that $U^* \lhd \lhd L$ and $U^* \supseteq U$. If the base field of the algebra is of characteristic zero, then by Theorem 7, U_* is maximal such that $U_* \lhd \lhd L$ and $U_* \subseteq U$. We observe that $U_* \subseteq U \subseteq U^*$.

The following results analogous to those for finite groups are then valid.

Lemma 7. *If U is any subalgebra of a Lie algebra L over a field of characteristic zero, then U_* is an ideal of U.*

Proof. Let b be any element of U. If U_1 is the minimal algebra such that $U_1 \supseteq U_*$ and $[U_1, b] \subseteq U_1$ then it follows from Theorem 8 that U_1 is subinvariant in L. Obviously $U_1 \subseteq U$ and therefore $U_1 \subseteq U_*$. It follows that $[U_*, b] \subseteq U_*$ for all $b \in U$ and therefore U_* is an ideal of U.

Theorem 13. *If U and V are subalgebras of a Lie algebra L over an arbitrary field and if $\{U, V\} = U + V$, then $\{U^*, V^*\} = U^* + V^*$.*

Proof. It will be sufficient to show that $U^* = U + U^{*\omega}$ and $V^* = V + V^{*\omega}$, since $U^{*\omega}$ and $V^{*\omega}$ are ideals of L. Obviously $U^* \supseteq U + U^{*\omega}$. On the other hand, $(U + U^{*\omega}) - U^{*\omega}$ is a subalgebra of the nilpotent algebra $U^* - U^{*\omega}$. Hence by Theorem 1 (g), $(U + U^{*\omega}) - U^{*\omega}$ is subinvariant in $U^* - U^{*\omega}$, which by Theorem 1 (e) is subinvariant in $L - U^{*\omega}$. Then by Theorem 1 (f), $U + U^{*\omega}$ is subinvariant in L and therefore $U + U^{*\omega} \supseteq U^*$. It follows that $U^* = U + U^{*\omega}$. Likewise $V^* = V + V^{*\omega}$ and the proof is complete.

The following example shows that $\{U, V\} = U + V$ need not imply that $\{U_*, V_*\} = U_* + V_*$. Let $L = S \oplus N$ where S is semisimple and N is the nilpotent algebra with basis x, y, z such that $[x, y] = z$, $[x, z] = [y, z] = 0$; and let N and S be ideals of L. Then if s is an element of S, consider the subalgebras $U = \{s, x\}$, $V = \{y, s - z\}$. Then $\{U, V\} = \{s, x, y, s - z\} = (x, y, z, s) = U + V$. But $U_* = (x)$, $V_* = (y)$; and $\{U_*, V_*\} = \{x, y\} = (x, y, z) \neq U_* + V_*$.

We now give a generalization of Theorem 7 to the case where one of the algebras need not be subinvariant.

Theorem 14. *Let A be subinvariant in a Lie algebra over a field of characteristic zero. Let V be any subalgebra of L, and let $K = \{A, V\}$. Then the composition factors from K^* to K_* are a part of those from V^* to V_*.*

Proof. By Lemma 7, K_* is an ideal of K. Therefore since $V \subseteq K$, $\{K_*, V\} = K_* + V$. Then by the preceding theorem $\{K_*, V^*\} = K_* + V^*$.

It follows that the composition factors from $K_* + V^*$ to K_* are the same as those from V^* to $V^* \cap K_*$. For by Theorem 2 the composition factors from V^* to $V^* \cap K_*$ occur those from $K_* + V^*$ to K_*. On the other hand if two algebras of a series from $K_* + V^*$ to K_* have the same intersections with V^*, then by their forms they must be the same algebras and hence there is a one-to-one correspondence between the composition factors of the two series, $K_* + V^*$ to K_*, and V^* to $V^* \cap K_*$.

Now $V \subseteq K$ and therefore $V_* \subseteq V^* \cap K_*$. Also $A \subseteq K_*$, $V \subseteq V^*$; consequently $K = \{A, V\} \subseteq \{K_*, V^*\} = K_* + V^*$ and $K^* \subseteq K_* + V^*$. We then have that $V_* \subseteq V^* \cap K_* \subseteq V^*$ and $K_* \subseteq K^* \subseteq K_* + V^*$ and it is easy to check that the theorem follows.

Now let k be any element of a Lie algebra L and denote the nil algebra determined by the derivation $\tilde{k}$ by K. We recall that k is called a *regular element* if $\dim K$ is minimal for all the elements k of L; and in this case K is called a *Cartan subalgebra.* The theory of these subalgebras is fundamental in the general structure theory of Lie algebras. The following result may therefore be of interest.

THEOREM 15. *If K is a Cartan subalgebra of a Lie algebra L over an arbitrary field, then* (1) $K^* = L$. (2) *K_* need not be an ideal of L.*[9]

Proof. Let K be the nil algebra of k. Then $L = K \oplus L_1$, where L_1 is the preserved space of $\tilde{k}$. By Lemma 6, $K^* \supseteq L_1$. Hence $K^* \supseteq L_1 + K = L$ and $K^* = L$, proving (1). To prove (2) we construct the algebra $A = \{a_1, a_2, b, k\}$ for which a partial multiplication table is defined to be $[a_1, a_2] = 0$, $[a_1, b] = 0$, $[a_1, k] = a_1$, $[a_2, b] = a_1$, $[a_2, k] = a_2$, $[b, k] = 0$; the other products are defined in the obvious way and it is easy to see that A is an algebra. The subalgebra $K = \{b, k\} = (b, k)$ is the nil algebra of $\tilde{k}$ and is a Cartan subalgebra. For otherwise there would be an element whose adjoint mapping determines a one-dimensional nil algebra and hence a preserved space of dimension three. But this is impossible since $A' = [A, A] = (a_1, a_2)$ is of dimension two. Therefore K is a Cartan subalgebra as asserted. Now K_* cannot contain k, else by Lemma 6 K_* would contain the preserved space of $\tilde{k}$, which is (a_1, a_2). On the other hand $B = (b) \lhd \{b, a_1\} \lhd \{b, a_1, a_2\} \lhd A$, and therefore $B = K_*$. But B is not an ideal of A, for $[b, a_2] = -a_1 \notin B$; and the proof is complete.

The above algebra A also illustrates that the centralizer of a subinvariant need not be subinvariant. For if Z is the centralizer of B, then $Z = (a_1, b, k)$. But $[a_2, k] = a_2$ and therefore Z is not subinvariant in A.

---

[9] Compare Wielandt's result [7], p. 229. If S is a Sylow subgroup of a finite group G, then S^* and S_* are normal subgroups of G.

8. The centralizer of a subinvariant subalgebra. The object of this and the following sections will be to prove the tower theorem for Lie algebras. The results depend only on those of Section 1 through 3 and are valid for an arbitrary base field.

We begin with two lemmas.

LEMMA 8. *If H is a nilpotent Lie algebra with center Q, and if $P \neq (0)$ is an ideal of H, then $(P \cap Q) \neq (0)$.*

Proof. For some n,

$$[\cdots[P, \overbrace{H], \cdots, H}^{n-1 \text{ times}}] \neq (0), [\cdots[P, \overbrace{H], \cdots, H}^{n \text{ times}}] = (0),$$

since H is nilpotent. Therefore $[\cdots[P, \overbrace{H], \cdots, H}^{n-1 \text{ times}}] \subseteq Q$. But $[\cdots[P, \overbrace{H], \cdots, H}^{n-1 \text{ times}}] \subseteq P$ since P is an ideal. Therefore $P \cap Q \neq (0)$ as was asserted.

LEMMA 9. *Let A be a Lie algebra over an arbitrary field and let $A^{\omega} = \bigcap_{k=1}^{\infty} A^k$. If Z is the centralizer of A^{ω} in A, and if Z is not contained in A^{ω}, then A has a center not equal to zero.*

Proof. A^{ω} is an ideal of A and consequently so also is Z. By Theorem 4, $A = A^{\omega} + H$, where H is a nilpotent algebra. Let $A_1 = Z + H$. Then A_1 is an algebra, since Z is an ideal and H is an algebra. Hence $A_1 = A_1{}^{\omega} + H_1$, where H_1 is a nilpotent algebra. But $A_1{}^{\omega} = (Z + H)^{\omega} \subseteq Z$ by the nilpotency of H and because Z is an ideal. If $Z \subseteq A_1{}^{\omega}$, then $Z \subseteq A^{\omega}$, since $A_1{}^{\omega} \subseteq A^{\omega}$. But this contradicts the hypothesis that $Z \not\subseteq A^{\omega}$. It follows that there is a $z_1 \varepsilon Z$, $z_1 \not\varepsilon A_1{}^{\omega}$. We shall write $z_1 = a_1{}^{\omega} + h_1$, where $a_1 \varepsilon A_1{}^{\omega}$ and $h_1 \varepsilon H_1$. This is permissible since $Z \subseteq A_1{}^{\omega} + H_1$. Now $h_1 \neq 0$, else $z_1 \varepsilon A_1{}^{\omega}$, contrary to the choice of z_1. Hence $h_1 = z_1 - a_1{}^{\omega} \varepsilon Z$, since $z_1 \varepsilon Z$ and $a_1{}^{\omega} \varepsilon A_1{}^{\omega} \subseteq Z$. It follows that $H_1 \cap Z \neq (0)$.

Now $H_1 \cap Z$ is an ideal of H_1, since Z is an ideal of A. H_1 has a center $Q \neq 0$, since H_1 is nilpotent. Consequently by the previous lemma $P = (H_1 \cap Z) \cap Q \neq 0$. Finally, P is in the center of A. For $A = A^{\omega} + H$, where $H \subseteq A_1^\omega + H_1$; and hence $H \subseteq A^\omega + H_1$, since $A_1^\omega \subseteq A^\omega$. Consequently $A = A^\omega + H_1$. But P is in the centralizer of A^ω since $P \subseteq Z$, and P is in the center of H_1 since $P \subseteq Q$. Therefore P is in the center of A, proving the lemma.

We can now prove the following theorem.

THEOREM 16. *Let L be a Lie algebra over an arbitrary field. If A is subinvariant in L and if A has centralizer* (0), *then the centralizer Z of $A^\omega = \bigcap_{k=1}^{\infty} A^k$ is contained in A^ω.*

Proof. $A \cap Z$ is contained in A^ω for if not, then by the previous lemma, A would have a center not equal to zero, contrary to hypothesis. Suppose Z is not contained in A^ω; then by the above remark Z is not contained in A. Now A^ω is an ideal of L and consequently so also is Z. It follows that $A + Z = K$ is an algebra, and $A \neq K$ since $Z \not\subseteq A$. Then since $A \lhd \lhd L$, $A \lhd A_1 \lhd \lhd K$, where $A_1 \neq A$. Let a_1 be an element of A_1, $a_1 \notin A$. By the form of K, $a_1 = a + z_1$, where $a \in A$, $z_1 \in Z$, and $z_1 \notin A$ since $a_1 \notin A$. Let $B = A + (z_1)$. B is an algebra since A is an ideal of A_1. Let $B^\omega = \bigcap_{k=1}^{\infty} B^k$. Then $B^\omega \overset{?}{=} A^\omega$. For $B^k \supseteq A^k$ for all k since $B \supset A$. On the other hand, $[A, z_1] \subseteq A \cap Z$ since A is an ideal of A_1 and Z is an ideal of L. But $A \cap Z \subseteq A^\omega$, and hence $B^k \subseteq A^k$ for all $k > 1$. Then since $z, \notin A^\omega$, by the previous lemma we know that B has a center and consequently A has a nonzero centralizer in L. But this is contrary to hypothesis and hence Z must be contained in A^ω, as the theorem asserts.

9. **The Tower theorem.**

Before we state the tower theorem it is convenient to have the following lemma.

LEMMA 10. *Let L be a Lie algebra with center zero, and let D_1 be the derivation algebra of L, D_0 the algebra of all inner derivations of L. Then* (1) *L is isomorphic to D_0*; (2) *D_0 is an ideal of D_1*; (3) *the centralizer of D_0 in D_1 is zero.*

Proof. Since L has center zero, it is evident that the representation of L by D_0 is an isomorphism, proving (1). To show that D_0 is an ideal of D_1, we let $\tilde{d} \in D_0$ correspond to $d \in L$; that is, for $x \in L$, $x\tilde{d} = [x, d]$. Then for arbitrary $\tilde{d}_1 \in D_1$, $x[\tilde{d}_1, \tilde{d}] = x(\tilde{d}_1\tilde{d} - \tilde{d}\tilde{d}_1) = [x\tilde{d}_1, d] - [x, d]\tilde{d}_1 = [x\tilde{d}_1, d] - [x\tilde{d}_1, d] - [x, d\tilde{d}_1] = -[x, d\tilde{d}_1]$. Hence $[\tilde{d}_1, \tilde{d}]$ is an inner derivation, which means that D_0 is an ideal of D_1, proving (2).

Finally, if $\tilde{d}_1$ is in the centralizer of D_0 in D_1, then $[\tilde{d}_1, \tilde{d}] = 0$ for all $\tilde{d} \varepsilon D_0$; whence by the above calculation $[x, d\tilde{d}_1] = 0$ for all x. But L has center zero. Therefore $d\tilde{d}_1 = 0$ for all $d \varepsilon L$. Hence $\tilde{d}_1 = 0$, proving the last statement of the lemma.

It follows from the lemma that if L has center zero, then $L \cong D_0$, $D(L)$ has no center, and $L \lhd D(L)$. Accordingly we can make the following definition.

Definition. If L is a Lie algebra with center zero, let $D_0 = L$, $D_1 = D(L)$, the derivation algebra of $L, \cdots, D_n = D(D_{n-1}), \cdots$, then the series of algebras $D_0, D_1, D_2, \cdots$ is called the *tower of derivation algebras* of L.

We wish to find a bound for the dimensions of the algebras D_k, $k = 1, 2, \cdots$, of the tower of derivation algebras of an arbitrary Lie algebra with center zero. The result is given in the following theorem.

THEOREM 17. *Let L be a Lie algebra over an arbitrary field and let the center of L be zero, so that $L = D_0$. Then $D_0, D_1, \cdots, D_n$, the tower of derivation algebras of L, ends finitely, and* $\dim D_n \leqq d + c$, *where d is the dimension of the derivation algebra of $L^\omega = \bigcap_{k=1}^{\infty} L^k$, and c is the dimension of the center of L^ω.*

Two lemmas are needed in the proof.

LEMMA 11. *If A is an ideal of L and Z is the centralizer of A in L, then* $\dim L \leqq \dim Z + d$, *where* $d = \dim D(A)$, *the derivation algebra of A.*

Proof. For $l \varepsilon L$, $[A, l] \subseteq A$, since A is an ideal of L. It follows that $\tilde{l} = \operatorname{ad} l$ acting in A is a derivation of A. Consequently there are at most d linearly independent derivations $l_1, \cdots, l_r$, where $r \leqq d$. Then the mapping of L onto the space $(l_1, \cdots, l_r)$ has Z for its kernel. Therefore, $\dim L = \dim Z + r$. But $r \leqq d$, and the lemma follows.

LEMMA 12. *If $A \lhd A_1 \lhd \lhd L$, if the centralizer of A in A_1 is zero, and if the centralizer of A_1 in L is zero, then the centralizer of A in L is zero.*

Proof. We will suppose on the contrary that A has centralizer $Z \neq (0)$ in L. Then $Z \cap A_1 = (0)$ by hypothesis. Consequently $P = \{A_1, Z\} \neq A_1$. Now $A \lhd A_1$ and $[A, Z] = (0)$. Therefore $A \lhd P$ and hence $Z \lhd P$. It follows that $P = A_1 + Z$. But $A_1 \lhd \lhd L$. Therefore $A_1 \lhd \lhd P \neq A$, and $A_1 \lhd A_2 \lhd \lhd P$, $A_2 \neq A_1$. Choose $a_2 \varepsilon A_2$, $a_2 \not\varepsilon A_1$. By the form of P, $a_2 = a_1 + z$, $a_1 \varepsilon A_1$, $z \varepsilon Z$, where $z \neq 0$ since $a_2 \not\varepsilon A_1$. Then $[z, A_1] \subseteq A_1 \cap Z$

since A_1 is an ideal of A_2 and Z an ideal of P. But $A_1 \cap Z = (0)$ and therefore $z \neq 0$ is in the centralizer of A_1, contradicting the hypothesis that A_1 had centralizer zero in L. Thus the lemma is established.

The proof of Theorem 17 will now be given. By Lemma 10, $L \cong D_0 \lhd D_1 \lhd D_2 \lhd \cdots \lhd D_n \lhd D_{n+1} \lhd \cdots$ where, for all l, D_{l-1} has zero centralizer in D_l. Suppose that D_{n-p} has zero centralizer in D_n. Then it follows from Lemma 12 that $D_{n-(p+1)}$ has zero centralizer in D_n. Since the statement is true for $p = 1$ by Lemma 10, it is true for all p; in particular, for $p = n$, whence $D_{n-n} = D_0 \cong L$ has zero centralizer in D_n. It follows from Theorem 16 that the centralizer of L^{ω} in D_n is contained in L^{ω} and hence has dimension c. Furthermore, L^{ω} is an ideal of D_n and hence by Lemma 11, $\dim D_n \leq d + c$, as the theorem asserts.

Since D_n has center zero for all n, the dimension of D_n is non-decreasing as a function of n, and it is immediate from the theorem that for some n all the derivations of D_n are inner. This result was announced by C. Chevalley [2] for Lie algebras of characteristic zero.

As a corollary of the Tower Theorem we have the following result.

THEOREM 18. *Let L be a Lie algebra with zero center over an arbitrary field. If L is isomorphic to an ideal K of the holomorph of L^{ω} such that K contains L^{ω}, then* $\dim D(L) = d + c$, *where again* $d = \dim D(L^{\omega})$ *and c is the dimension of the center of L^{ω}. It follows that $D(L)$ has only inner derivations.*

Proof. We recall that the holomorph of L^{ω} is the semi-direct sum $H = L^{\omega} + D(L^{\omega})$, where for $l \varepsilon L^{\omega}$, $\tilde{d} \varepsilon D(L^{\omega})$, $[l, \tilde{d}]$ is defined to be $l\tilde{d}$. It is then obvious that L^{ω} is an ideal of H and hence of K. We have thus that $L^{\omega} \underset{=}{\lhd} K \underset{=}{\lhd} H$.

Now there are d linearly independent elements of H which determine linearly independent derivations of L^{ω} and hence of K. Furthermore, there are c linearly independent elements of H which are in the center of L^{ω} and hence are in K. But K is isomorphic to L and hence has no center. Therefore these c elements determine linearly independent derivations of K which are independent of the d elements above since they map L^{ω} into zero. Hence $\dim D(K) \geq d + c$ and consequently $\dim D(L) \geq d + c$. But from the Tower Theorem we know $\dim D(L) \leqq d + c$, and hence $\dim D(L) = d + c$, as asserted. Again it follows from the Tower Theorem that all the derivations of $D(L)$ are inner, and the proof is complete.

For the particular case where $L = L'$, the above result was proved by Zassenhaus.[10]

The following special case of the preceding theorem may also be of interest.

COROLLARY. *All the derivations of the holomorph of an Abelian Lie algebra are inner.*

Proof. Let A be an Abelian Lie algebra. Let $H = A + D(A)$ be the holomorph of A. Consider the subalgebra $L = A + (\tilde{j})$, where $\tilde{j}$ is the identity derivation of A; that is, $a\tilde{j} = a$ for all $a \varepsilon A$. Then L has no center. Furthermore it is easy to verify that $[\tilde{j}, \tilde{d}] = 0$ for all $\tilde{d} \varepsilon D(A)$, and hence L is an ideal of H. Finally, we note that $L^{\omega} = A$ and hence the previous theorem applies. Therefore $\dim D(L) = d + c$, where $d = \dim D(A)$, and $c = \dim A$ since A is Abelian. But $\dim H = d + c$, and since the centralizer of L in H is (0), it follows that $D(L) \cong H$, and the corollary follows.

LOUISIANA STATE UNIVERSITY.

BIBLIOGRAPHY.

[1] E. Cartan, *Sur la structure des groupes de transformation, finis et continus*, Paris, Libraire Vuibert (1933).

[2] C. Chevalley, "On groups of automorphism of Lie groups," *Proceedings of the National Academy of Sciences*, vol. 30 (1944), pp. 274-275.

[3] G. Hochschild, "Lie algebras and differentiations in rings of power series," *American Journal of Mathematics*, vol. 72 (1950), pp. 58-80.

[4] N. Jacobson, "Rational methods in the theory of Lie algebras," *Annals of Mathematics*, vol. 36 (1935), pp. 875-881.

[5] N. Jacobson, *The theory of rings*, American Mathematical Society, New York (1943).

[6] H. Weyl, "Theorie der Darstellung kontinuierlicher halb-einfacher Gruppen durch lineare Transformationen. II," *Mathematische Zeitschrift*, vol. 24 (1926), 328-376.

[7] H. Wielandt, "Eine Verallgemeinerung der invarianten Untergruppen," *Mathematische Zeitschrift*, vol. 45 (1939), pp. 209-244.

[8] H. Zassenhaus, "Über liesche Ringe mit Primzahlcharakteristik," *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität*, vol. 13 (1940), pp. 1-100.

---

[10] Zassenhaus [8], p. 52.

HOMOMORPHISMS OF A SEMIGROUP ONTO A GROUP.*

By R. R. Stoll.

Among the homomorphic images of a semigroup S ($=$ a set closed with respect to an associative binary operation) there is at least one group, namely the unit group I. The question arises as to the existence of maximal group images of some sort for a given S. The answer is trivial if S contains a zero element, i. e. an element 0 such that $0x = x0 = 0$ for all $x \varepsilon S$, since then I is the only group image. In such cases the question has a broader scope if asked in connection with images of the form, a group with a zero element adjoined. These images exist if and only if S contain a prime ideal, according to Theorem 1.

In the present paper the notions of a maximal group image and maximal group with zero image are formulated for a semigroup S and the construction of such images is discussed in terms of "normal" subsystems (generalizations of a normal subgroup in the group case) of S, using a method devised by Dubreil [2]. A set of independent defining conditions for these normal subsystems is obtained. Finally, several well known semigroups which have unique maximal images of the types under considerations are discussed.

The most interesting examples are two classes of semigroups (regular sets of partial transformations of a set, and completely simple semigroups without zero) studied by Rees in [7] and [5] respectively, and one (semigroups having zeroid elements) studied by Clifford and Miller in [1].

For the first class, the group image discussed by Rees is shown to be maximal. For the second class, the maximal group image can be described in terms of a homomorphic image of the basis group for the semigroup. For the third class, the group of zeroid elements is found to be the maximal group image.

1. **Group with zero images of a semigroup.** The symbol G^* will be used to denote the group G with a zero element adjoined. In order to characterize those semigroups which admit some G^* as image, a definition is needed. By an ideal of a semigroup S is meant a subset P such that SP, $PS \subseteq P$. If, moreover, $ab \varepsilon P$ implies that at least one factor is in P, it is a prime ideal.

---

* Received March 8, 1950; presented to the American Mathematical Society, September 7, 1948.

THEOREM 1. *A semigroup S admits a group with zero as image if and only if it contains a prime ideal.*

Proof. Given $\alpha: S \rightarrow G^*$, it is easily verified that $\alpha^{-1}(0)$, where 0 is the zero element of G^*, is a prime ideal. Conversely, if S contains the prime ideal P, S can be mapped onto I^*, where I is the unit group, as follows: $S - P \rightarrow I$, $P \rightarrow 0$. •

In view of this result, we restrict our attention in this section to semigroups S with a prime ideal. A homomorphism of S onto G^* is characterized by a pair (G^*, α) where α is the mapping function. For these pairs we make

Definition 1. $(G_1{}^*, \alpha_1) = (G_2{}^*, \alpha_2)$ if and only if $P(\alpha_1) = P(\alpha_2)$ where $P(\alpha_i)$ is the partition of S defined by the homomorphism α_i.

Equality of the above pair implies that $G_1{}^*$ and $G_2{}^*$ are isomorphic. As a representative of a class of equal pairs we may take the group with zero formed by the blocks of the common partition together with the homomorphism of inclusion. Henceforth (G^*, α) will denote such a pair.

In order to determine the possible distinct pairs (G^*, α) accompanying a given S, we use Dubreil's results [2] which we now describe. His basic notion is that of the right quotient of a complex $H \subset S$ by an element $a \in S$:

$$(H:a) = \{x \in S \mid ax \in H\}.$$

The principal right equivalence $\mathcal{E}_H$ induced in S by H is defined by $a \equiv b (\mathrm{mod}\, \mathcal{E}_H)$ if and only if $(H:a) = (H:b)$. The right residue R_H is the set of all $x \in S$ for which $(H:x)$ is empty; R_H is a class mod $\mathcal{E}_H$. The definitions of the principal left equivalence ${}_H\mathcal{E}$ and left residue ${}_HR$ are clear. In the event $\mathcal{E}_H = {}_H\mathcal{E}$ and $R_H = {}_HR$, i. e. H is symmetric, the common equivalence relation $\mathcal{E}$ is a congruence relation, hence defines a homomorphic image $S/\mathcal{E}$ of S where the correspondence is that of inclusion. Those H's for which simultaneously $S/\mathcal{E}$ is a group with zero and with which any given homomorphism (G^*, α) can be duplicated, Dubreil calls the pseudo-normal and unitary subsemigroups N^* of S. His results in this connection are summarized in

THEOREM 2. *The correspondence $(G^*, \alpha) \rightarrow N^* = \alpha^{-1}(e)$, where e is the unit element of G, is a one-one correspondence between the set $\mathfrak{G}$ of distinct elements (G^*, α) and the set $\mathfrak{N}$ of pseudo-normal and unitary subsemigroups N^* of S. Moreover, in the correspondence $\alpha: S \rightarrow G^*$, the residue R (a prime ideal) of N^* is $\alpha^{-1}(0)$ where 0 is the zero element of G^*.*

For the complexes N^* it is possible to deduce from Dubreil's paper the following set of independent defining conditions: (i) N^* is a subsemigroup, (ii) $a, b \varepsilon S$, and $ab \varepsilon N^*$ imply $ba \varepsilon N^*$, (iii) $a, b \varepsilon S$ and a and ab in N^* imply $b \varepsilon N^*$, (iv) the residue R_{N^*} is non-empty. Since the conditions (i)-(iii) imply that N^* is symmetric it is permissible to speak of the residue in (iv).

THEOREM 3. *The condition* (ii) *and* (iii) *above on a subsemigroup* N^* *of* S *are equivalent to the condition*

(v) $a, b, c \varepsilon S$ *and* abc *and* ac *in* N^* *imply* $b \varepsilon N^*$.

Proof. Assume that (v) holds. Then $ab \varepsilon N^*$ implies $(ab)^2 = a(ba)b \varepsilon N^*$, implies $ba \varepsilon N^*$, so that (ii) holds. Next, $a, ab \varepsilon N^*$ imply $aba, a^2 \varepsilon N^*$, hence $b \varepsilon N^*$ and (iii) holds. Conversely, if (ii) and (iii) hold, $abc, ac \varepsilon N^*$ imply $bca, ca \varepsilon N^*$, hence $b \varepsilon N^*$.

Definition 2. In $\mathfrak{G}$, call $(G_1^*, \alpha_1) < (G_2^*, \alpha_2)$ if and only if the partition $P(\alpha_2)$ is a refinement of $P(\alpha_1)$.

THEOREM 4. $(G_1^*, \alpha_1) < (G_2^*, \alpha_2)$ *if and only if* $N_1^* \supseteq N_2^*$ *and* $R_1 = R_2$, *where* R_i *is the residue of* N_i^*, *which in turn is the correspondent of* (G_i^*, α_i) *in accordance with Theorem* 2.

Proof. If $(G_1^*, \alpha_1) < (G_2^*, \alpha_2)$ then $N_2^* \subseteq N_1^*$ and $R_2 \subseteq R_1$. But by the definition of the residue, $x \varepsilon R_1$ implies $x \varepsilon R_2$, so that $R_1 \subseteq R_2$ and equality follows.

Conversely, if $N_2^* \subseteq N_1^*$ and $R_1 = R_2$ we shall prove that $a \equiv b (\operatorname{mod} \mathcal{E}_{N_2^*})$ implies $a \equiv b (\operatorname{mod} \mathcal{E}_{N_1^*})$. By assumption this is true for elements in the residue, so suppose that $(N_2^* : a) \neq O$. If $c \varepsilon (N_2^* : a)$ then ac and bc lie in N_2^*, and hence in N_1^*. If now $ax \varepsilon N_1^*$, we use the foregoing conditions (i), (v), (iii) for pseudo-normal and unitary subsemigroups in turn, to conclude that since $ax \cdot bc = a(xb)c$ and ac lie in N, so does xb, hence bx. Thus $(N_1^* : a) \subseteq (N_1^* : b)$. The reverse inequality is established similarly and equality holds; this means $a \equiv b (\operatorname{mod} \mathcal{E}_{N_1^*})$.

The set $\mathfrak{N}$ is the sum of the disjoint subsets $\mathfrak{N}_R = \{N^* \varepsilon \mathfrak{N} \mid R \text{ is the residue of } N^*\}$. This decomposition of $\mathfrak{N}$ defines a decomposition of $\mathfrak{G}$ into disjoint classes $\mathfrak{G}_R$. If $\mathfrak{N}_R$ is partially ordered by set inclusion and $\mathfrak{G}_R$ is partially ordered by Definition 2, Theorem 4 implies

THEOREM 5. *The one-one correspondence* $(G^*, \alpha) \rightarrow N^*$ *of the partially ordered sets* $\mathfrak{G}_R$ *and* $\mathfrak{N}_R$ *is an anti-isomorphism.*

Definition 3. By a maximal image of S of the type, a group with zero element adjoined, is meant a maximal element in $\mathfrak{S}$ when partially ordered by Definition 2.

In view of the disjunction of $\mathfrak{S}$ into the classes $\mathfrak{S}_R$ and Theorems 4 and 5, the problem of finding maximal (G^*, α)'s is that of finding minimal N^*'s in $\mathfrak{N}_R$ for each of the residues R, i. e. each of the prime ideals in S.

2. Group images of a semigroup. The preceding section can be modified to the case of group images of a semigroup S (with or without a zero element), by omitting *'s and the notion of residues. The complexes N which generate group images of S, Dubreil calls the normal and unitary subsemigroups of S. They are characterized by (i), (ii), (iii) and $\overline{\text{(iv)}}$ (the negation of (iv)), or alternatively by (i), $\overline{\text{(iv)}}$ and (v). The latter is a set of independent defining conditions which can be extxracted from the redundant set devised by Levi [3, 4] for defining his "complete and normal subsemigroups" in connection with the problem of determining all group images of a semigroup.

3. Examples of maximal group with zero images. We now obtain the maximal images of the form, a group with zero, for several semigroups.

Example 1. The semigroup M of all $n \times n$ matrices over a field. It is a routine matter to conclude that the only ideals of M, apart from M, are the sets M_r consisting of all matrices of rank $\leqq r$, $r = 0, 1, \cdots, n-1$ and that among these only M_{n-1} is a prime ideal. Associated with this unique residue is a unique minimal pseudo-normal and unitary subsemigroup, namely the unit matrix I. The corresponding maximal group with zero image is, of course, the group of non-singular matrices with zero adjoined. The only interest in this obvious result is the uniqueness of the maximal image.

Example 2. The semigroup of all correspondences of a finite set C onto subsets of C. The map of C onto the empty subset is the zero element. An analysis along the lines of Example 1 yields the group of al permutations of C together with zero as the unique maximal image. The same conclusion holds for

Example 3. The semigroup of all partial transformation of a finite set C. Rees [7] has defined a partial transformation $\alpha: x \rightarrow x\alpha$ of a set C as a one-one correspondence of a subset C onto a subset of C. If $A(\alpha)$ denotes the set of elements for which α is defined and $V(\alpha)$ the set of image elements, the product $\alpha\beta$ of two partial transformations is defined by iteration over $V(\alpha) \cap A(\beta)$. The zero element ζ is the partial transformation with both $V(\zeta)$ and $A(\zeta)$ the empty set; $\alpha\beta = \zeta$ if and only if $V(\alpha) \cap A(\beta)$ is empty.

4. Examples of maximal group images.

Example 4. A regular set of partial transformations of a set C. This is a motion introduced by Rees [7]. He assumes that the product of the transformations α, β is defined only if $V(\alpha) \cap A(\beta)$ is non-empty and calls a set Σ of partial transformations regular if (a) $\alpha, \beta \varepsilon \Sigma$ imply $\alpha\beta$ is defined and in Σ, and (b) $\alpha \varepsilon \Sigma$ implies $\alpha^{-1} \varepsilon \Sigma$, where α^{-1} is the transformation $x\alpha \rightarrow x$. A regular set Σ is a semigroup which may or may not have a zero element. Rees obtains a group image for Σ as follows. If for $\alpha, \alpha' \varepsilon \Sigma$, $A(\alpha') \subseteq A(\alpha)$ and for all $x \varepsilon A(\alpha')$, $x\alpha' = x\alpha$, α' is a subtransformation of α. The relation $\mathfrak{R} : \alpha \equiv \beta (\text{mod } \mathfrak{R})$ if and only if α and β have a common subtransformation in Σ is a congruence relation over Σ and the quotient $\Sigma/\mathfrak{R}$ of Σ by $\mathfrak{R}$ is a group homomorphic to Σ.

We now show that $\Sigma/\mathfrak{R}$ is the unique maximal group image of Σ. If $\alpha \varepsilon \Sigma$, so is the identity transformation $\epsilon = \alpha\alpha^{-1}$. Thus the unit residue class U mod $\mathfrak{R}$ consists of all elements having a subtransformation in common with any one fixed identity transformation. If on the other hand N is any normal and unitary subsemigroup in Σ, it satisfies (i), (ii), (iii), (iv) and (v) of § 1. From these conditions we conclude:

(1) If $\alpha \varepsilon N$, so is α^{-1}; thus N contains an identity $\epsilon = \alpha\alpha^{-1}$. For along with $\alpha = \alpha\alpha^{-1}\alpha$, N contains $\alpha\alpha$ by (i), hence α^{-1} by (v).

(2) If β has $\gamma \varepsilon N$ as a subtransformation then $\beta \varepsilon N$. For $\gamma\gamma^{-1}$, $(\gamma\gamma^{-1})\beta = \gamma \varepsilon N$ imply $\beta \varepsilon N$ by (iii).

(3) If β and the identity ϵ have a common subtransformation $\gamma \varepsilon \Sigma$ then $\beta \varepsilon N$. For γ, as a subtransformation of ϵ is idempotent and hence in N. The assertion then follows from (2).

Thus every normal and unitary N contains the above U and the uniqueness of U as the minimal normal and unitary subsemigroup is established.

Example 5. A completely simple semigroup without zero. Rees [5, 6] has defined such a semigroup S by the conditions, (a) the only ideal in S is S itself, (b) S contains a primitive idempotent, i. e. an idempotent e such that there is no idempotent $f \neq e$ for which $ef = fe = f$. Such a semigroup can be described as the set of all triplets

$$(a; \alpha, \beta) \qquad a \varepsilon H, \qquad \alpha \varepsilon A, \qquad \beta \varepsilon B,$$

where H is a group, the basis group for S, and A and B are two sets of indices. To each pair α, β of indices ($\alpha \varepsilon A, \beta \varepsilon B$) there is assigned some fixed element $p_{\beta\alpha} \varepsilon H$ and multiplication is defined by $(a; \alpha, \beta)(b; \alpha', \beta')$

$= (ap_{\beta\alpha'}b\,;\alpha,\beta')$. We shall call the $p_{\beta\alpha}$ the constants of S. It is possible to alter these constants somewhat without changing S. For example, if 1 is taken to be a member of both sets A, B of indices, we may choose $p_{\alpha 1} = p_{1\beta} = e$, the unit element in H, for all $\alpha \varepsilon A$, $\beta \varepsilon B$. We assume this is done. The idempotents of S are the elements $e_{\alpha\beta} = (p^{-1}{}_{\beta\alpha}\,;\alpha,\beta)$. The set of elements $(a\,;\alpha,\beta)$ where a varies over H and α, β are fixed, form a group isomorphic to H, so that S is the sum of disjoint groups, each isomorphic to H.

THEOREM 6. *Let S denote a completely simple semigroup without zero, with basis group H, index sets A, B and constants $p_{\beta\alpha}$. Let P denote the (unique) minimal normal subgroup of H containing the $p_{\beta\alpha}$. Then S contains a unique minimal normal and unitary subsemigroup $N = \sum_{\alpha\varepsilon A,\beta\varepsilon B} (P\,;\alpha,\beta)$, where $(P\,;\alpha,\beta)$ is the totality of elements $(p\,;\alpha,\beta)$ with $p \varepsilon P$. Thus S has a unique maximal group image. This group is isomorphic to H/P.*

Proof. We first show that a normal and unitary subsemigroup M of S must contain N, using the conditions in § 1. Since M must contain the idempotents $e_{\alpha\beta}$ of S, it contains $e_{1\beta}e_{\alpha 1} = (p_{\beta\alpha}\,;1,1)$. Moreover $(p_{\beta\alpha}\,;1,1)(p_{\beta\alpha}{}^{-1}\,;1,1) = e_{11} \varepsilon M$, hence by (iii) $(p_{\beta\alpha}{}^{-1}\,;1,1) \varepsilon M$. The equations

$$(x^{-1};1,1)(x\,;1,1) = e_{11} \varepsilon M, \quad (x^{-1};1,1)(xp_{\beta\alpha}x^{-1};1,1)(x\,;1,1) \varepsilon M$$

and (v) imply that $(xp_{\beta\alpha}x^{-1};1,1) \varepsilon M$; similarly for $(xp_{\beta\alpha}{}^{-1}x^{-1};1,1)$. Since M is closed, it follows that $(P\,;1,1) \subset M$. Finally, the equations

$$e_{11}e_{11} = e_{11} \varepsilon M,\ e_{11}\ (p\,;\alpha,\beta)e_{11} = (p\,;1,1) \varepsilon M \text{ for } p \varepsilon P,$$

and (v) imply that $(p\,;\alpha,\beta) \varepsilon M$. Thus $M \supset N$.

To show that N is normal and unitary it suffices to verify that (i), (iv) and (v) hold. Clearly N is closed. As for (iv), multiplication on the right of an element $(a\,;\alpha,\beta) \varepsilon S$ by $((p_{\alpha\delta}ap_{\beta\gamma})^{-1};\gamma,\delta)$ gives $e_{\alpha\delta} \varepsilon N$. Finally, suppose that the products $(a\,;\alpha,\beta)(b\,;\gamma,\delta)(c\,;\lambda,\mu)$ and $(a\,;\alpha,\beta)(c\,;\lambda,\mu)$ are in N. These products then have the form $(p_1\,;\alpha,\mu)$ and $(p_2\,;\alpha,\mu)$ respectively, where $p_i \varepsilon P$. If one calculates $a^{-1}p_1p_2{}^{-1}a$ it is immediately seen that $b \varepsilon P$, hence that $(b\,;\gamma,\delta) \varepsilon N$. Thus (v) holds.

The proof of the isomorphism of $S/\mathcal{E}_N$ and H/P is immediate. One finds that the cosets of P in H describe the equivalence classes in $S \bmod \mathcal{E}_N$, i. e. an equivalence class in $S \bmod \mathcal{E}_N$ has the form $\sum_{\alpha\varepsilon A,\beta\varepsilon B} (hP\,;\alpha,\beta)$. The correspondence of this class with the coset hP of P in H establishes the isomorphism.

Example 6. Semigroups having zeroid elements. Clifford and Miller [1] have defined an element u of a semigroup S as a zeroid element if, for each element a of S, there exist x and y in S such that $ax = ya = u$. If the set U of all zeroid elements in S is non-empty, U is a group and a homomorph of S under the correspondence $\zeta : x \rightarrow xz$ where z is the unit element of U. The set $J = \zeta^{-1}(z)$ has been characterized by Clifford and Miller as the set of all elements in S for which z is a zero element.

We shall prove that U is the unique maximal group image of S by verifying that if N is any normal and unitary subsemigroup in S then $N \supseteq J$. By (iv), for an element $a \varepsilon J$, there exists an $x \varepsilon S$ such that $ax \varepsilon N$. As an idempotent, $z \varepsilon N$, hence $zax = zx \varepsilon N$ by (i). But z and zx in N imply $x \varepsilon N$ by (iii). Finally, x and $ax \varepsilon N$ imply $a \varepsilon N$ by (iii). Thus $J \subseteq N$. Moreover the map $S \rightarrow S/\mathcal{E}_J$ (where the homomorphism is that of inclusion) is the map $\zeta : x \rightarrow xz$.

5. Remarks. In contrast to the foregoing examples, the additive semigroup P of positive integers is an instance of a semigroup having no maximal group images. For clearly the group images of P are precisely the totality of finite multiplicative cyclic groups and among these there is no maximal group.

The problem of finding useful reformulations of the condition obtained earlier (namely the existence of a minimal normal and unitary subsemigroup) for the existence of a maximal group image appears to be difficult. The same is true for the problem of determining when a given semigroup has a unique maximal group image. These matters will be discussed elsewhere.

LEHIGH UNIVERSITY.

REFERENCES.

[1] A. H. Clifford and D. D. Miller, "Semigroups having zeroid elements," *American Journal of Mathematics*, vol. 70 (1948), pp. 117-125.

[2] Paul Dubreil, "Contribution à la théorie des demi-groupes," *Mémoires de l'Académie des Sciences*, ser. 2, vol. 63 (1941), pp. 1-52.

[3] F. W. Levi, "On semigroups," *Bulletin of the Calcutta Mathematical Society*, vol. 36 (1944), pp. 141-146.

[4] ———, "On semigroups II," *ibid.*, vol. 38 (1946), pp. 123-124.

[5] D. Rees, "On semigroups," *Proceedings of the Cambridge Philosophical Society*, vol. 36 (1940), pp. 387-400.

[6] ———, "Note on semigroups," *ibid.*, vol. 37 (1941), pp. 434-435.

[7] ———, "On the group of a set of partial transformations," *Journal of the London Mathematical Society*, vol. 22 (1947), pp. 281-284.

A NOTE ON A FUNCTIONAL EQUATION.*

By H. P. THIELMAN.

We consider the functional equation

$$\phi[F(x,y)] = \phi(x) + \phi(y) \tag{1}$$

and determine the conditions on a polynomial $F(x,y)$ in order that (1) can have a continuous strictly monotone (say increasing) solution. If $F(x,y)$ is $x+y$ or xy, we have the classical cases of Cauchy's functional equations. The cases when $F(x,y)$ is a general linear function,[1] and also the case when it is $x+y+nxy$, have been treated.[2] In this paper the case when $F(x,y)$ is a polynomial of degree greater than unity is settled, as follows:

THEOREM. *If $F(x,y)$ is a polynomial of degree greater than unity for which* (1) *has a continuous monotone increasing solution, then $F(x,y)$ must be of the form*

$$F(x,y) = \{(ax+b)(ay+b) - b\}/a, \qquad a > 0.$$

Proof. By hypothesis, equation (1) has a strictly monotone increasing, continuous solution $\phi(x)$. Hence the inverse $\phi^{-1}(x)$ will exist and we can write (1) as

$$F(x,y) = \phi^{-1}[\phi(x) + \phi(y)]. \tag{2}$$

In order to simplify the notation we shall denote $F(x,y)$ by $x \circ y$. This can be interpreted as an operation which, because of (2), has the following properties: (i) $x \circ y = y \circ x$, commutativity; (ii) $x \circ (y \circ z) = (x \circ y) \circ z$, associativity; (iii) $x' \circ y > x \circ y$ if $x' > x$, strict monotony; (iv) continuity. [From (iii) and (i) it follows that $x \circ y' > x \circ y$ if $y' > y$, $x \circ x < x \circ y < y \circ y$ if $x < y$. Aczél has shown that any operation satisfying the above conditions can be written in form (2)].

Let $x \circ y$ be a polynomial of degree n in x, and m in y. Then, because of (i), $n = m$. Because of (ii), $(x \circ y) \circ z = x \circ (y \circ z)$. The left hand side of this equation is a polynomial of degree n in z, the right hand side a

---

* Received July 8, 1949.

[1] J. Aczél, *Comm. Math. Helvetici*, vol. 21 (1948), pp. 247-252.

[2] H. P. Thielman, *American Mathematical Monthly*, vol. 56 (1949), pp. 452-457.

polynomial of degree n^2 in z. Hence $n=1$. Thus $x \circ y$ is a symmetric polynomial of degree 2 in x and y combined. It can be written as $x \circ y = Bxy + C(x+y) + D$.

Let us denote $x \circ x$, which is of degree 2, by $P_2(x)$, and $x \circ (x \circ x)$, which is of degree 3, by $P_3(x)$. By (ii), $[x \circ x] \circ [x \circ (x \circ x)] = [x \circ (x \circ x)] \circ (x \circ x)$, which states that $P_2(x)$ and $P_3(x)$ are commutative. But it is easily proved, by equating coefficients, that if a quadratic polynomial commutes with a cubic one, then it must be of the form $\lambda^{-1}[\lambda^2(x)]$ or $\lambda^{-1}[\cos 2 \operatorname{arc} \cos \lambda(x)]$, where $\lambda(x)$ is linear, say $ax+b$, and $\lambda^{-1}(x)$ is its inverse. Thus either

a) $x \circ x = Bx^2 + 2Cx + D \equiv \{(ax+b)^2 - b\}/a$

or

b) $x \circ x = Bx^2 + 2Cx + D \equiv \{2(ax+b)^2 - 1 - b\}/a.$

We first consider the identity a). It implies that $B=a$, $C=b$, $D=(b^2-b)/a$, and $F(x,y) = \{(ax+b)(ay+b) - b\}/a$. It is easily seen that conditions (i), (ii), (iii), (iv) are satisfied if $x > -b/a$, $y > -b/a$, $a > 0$. If we let $\phi(x) = \log(ax+b)$, then $F(x,y) = \phi^{-1}[\phi(x) + \phi(y)]$, or

$$\phi[\{(ax+b)(ay+b) - b\}/a = \phi(x) + \phi(y).$$

[By a method similar to the one used *loc. cit.*,[2] it can be shown that every continuous solution of this last functional equation is of the form $k \log(ax+b)$, where k is an arbitrary constant.]

We next consider the identity b). It can be written as $Bx^2 + 2Cx + D \equiv 2ax^2 + 4bx + (2b^2 - 1 - b)/a$, from which it follows that

$$x \circ y = F(x,y) = 2axy + 2b(x+y) + (2b^2 - 1 - b)/a.$$

Condition (ii) is not satisfied for $(x \circ x) \circ z \neq x \circ (x \circ z)$, as can be easily verified. Hence this case is eliminated and the theorem is proved.

The theorem could be stated in the terminology of the literature[3] dealing with operations as follows: Every rational, integral, associative and commutative operation which is of degree greater than unity is of the form $\lambda^{-1}[\lambda(x)\lambda(y)]$, where $\lambda(x)$ is a linear function.

COROLLARY. *Every associative and commutative operation $x \circ y$ which for $x = y$ reduces to a polynomial of degree greater than unity must be such that either $x \circ x = \lambda^{-1}[\lambda^2(x)]$ or $x \circ x = \lambda^{-1}[\cos 2 \operatorname{arc} \cos \lambda(x)]$.*

The proof of this corollary was given in the process of proving the theorem.

---

[3] J. Aczél, *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 392-400.

That there exist algebraic associative, commutative and increasing operations which are not integral is shown by the example

$$x \circ y = \{(ax+b)(ay+b) - [1-(ax+b)^2]^{\frac{1}{2}}[1-(ay+b)^2]^{\frac{1}{2}} - b\}/a,$$

which can be written as

$$x \circ y = \{\cos[\operatorname{arc\,cos}(ax+b) + \operatorname{arc\,cos}(ay+b)] - \mathrm{b}\}/a.$$

In the interval $-b/a \leqq x, y \leqq (1-b)/a$, $(a > 0)$, this function satisfies all the conditions (i), (ii), (iii), (iv), and it also illustrates the corollary. Obviously this last function can serve in the functional equation (1), whose continuous solutions then are all of the form $\phi(x) = k \operatorname{arc\,cos}(ax+b)$, where k is any constant. But since $x \circ y$ in this case is not a polynomial we shall not concern ourselves with it in this paper. It might be of interest to note that every associative and commutative operation $x \circ y$ which for $x = y$ reduces to a polynomial of degree greater than unity will generate an "entire set of commutative polynomials"[4] which must be either linear transforms of powers of x or of the trigonometric polynomials $\cos n \operatorname{arc\,cos} x$.[5] By a linear transform of a function $f(x)$ is meant the function $\lambda^{-1}[f(\lambda(x))]$, where $\lambda(x)$ is any linear function $ax+b$, $a \neq 0$. To generate these polynomials we need to iterate $x \circ x$ with itself. It should be mentioned that if we had replaced "monotone increasing" by "monotone decreasing," the results would still be true but for intervals different from those indicated above.

IOWA STATE COLLEGE.

---

[4] B. Vinograde, *American Mathematical Monthly*, vol. 56 (1949), pp. 377-379.

[5] This follows from the results of J. F. Ritt, *Transactions of the American Mathematical Society*, vol. 25 (1923), pp. 339-448.

CALCU
CENTRAL LIBRARY
UNIVERSI

A THEORY OF DEGREE OF MAPPING BASED ON INFINITESIMAL ANALYSIS.*

By MITIO NAGUMO.

---

1. **Introduction.** This paper establishes a theory of degree of mapping for open sets in a Euclidean space of finite dimension, based on the theory of infinitesimal analysis, which is free from the notion of simplicial mapping. Although the results are not new, I hope in this way to make it possible to incorporate the theory of degree of mapping into a course in infinitesimal analysis.

A mapping $x' = f(x) = \{f_i(x) \mid i = 1, \cdots, m\}$ $(x = (x_1, \cdots, x_m))$ of a bounded open set G of an m-dimensional Euclidean space E^m into E^m is called *regular* on G, if each $f_i(x)$ is continuously differentiable on the closure $\bar{G}$ of G. We write $\partial x_i f$ or $\partial_i f$ for $\partial f/\partial x_i$, and by $D(f_1, \cdots, f_m/x_1, \cdots, x_m)$, or more concisely by $D(f/x)$, we denote the functional determinant $\det(\partial_i f_j)_{i,j=1,\ldots,m}$. We call the point x, for which $D(f/x)$ vanishes, a *critical point* of f; and the image of the set of all critical points (of f) by f will be called the *crease* of f on G. The expression $a \bar{\varepsilon} A$ means "a does not belong to A."

Now let a be a point of E^m which lies neither on $f(\bar{G} - G)$ (the image of the boundary of G) nor on the crease of f. Then, by a theorem on implicit functions, there exist at most a finite number of points x in G such that $f(x) = a$. Let the number of those points for which $D(f/x) > 0$ be p, and the number of those for which $D(f/x) < 0$ be q. Then we call the integer $p - q$ the *degrees of mapping* of G at a by f, and denote it by

$$A[a, G, f] \, (= p - q).$$

We have the following properties of the degree of mapping:

i) *If f is the identical mapping $f(x) = x$, then*

$A[a, G, f] = 1$ *when* $a \varepsilon G$, $A[a, G, f] = 0$ *when* $a \bar{\varepsilon} \bar{G}$.

ii) *If $A[a, G, f] \neq 0$ and has a meaning, then there exists a point $x \varepsilon G$, such that $f(x) = a$.*

---

* Received March 6, 1950; revised September 4, 1950.

Essentially the same treatise by the author was published in a brochure in the Japanese language with examples and applications, in which (following Leray and Schauder) the theory is extended to a class of mappings in Banach space.

iii) *If* G *is divided into open sets* G_i $(i=1,\cdots,k)$ *without common points, i. e.,* $G\supset G_1\cup G_2\cup\cdots\cup G_k$, $\bar{G}=\bar{G}_1\cup\bar{G}_2\cup\cdots\cup\bar{G}_k$ *and* $G_i\cap G_j=0$ *(empty set)* $(i\neq j)$, *then*

$$A[a,G,f]=\sum_{i=1}^{k}A[a,G_i,f],$$

if every term has a meaning.

iv) *If* $f_t(x)$ *is a continuous function of* (t,x) *for* $0\leqq t\leqq 1$, $x\in\bar{G}$, *and* $a(t)\,(\in E^m)$ *is continuous and* $a(t)\bar{\in} f_t(\bar{G}-G)$ *for all* t *in* $0\leqq t\leqq 1$, *then* $A[a(t),G,f_t]$ *is constant for* $0\leqq t\leqq 1$.

The propositions i), ii), and iii) are obvious; but for the proof of iv) we need considerable preparation. In § 2 we shall give auxiliary theorems for the proof of iv) which will be carried out in § 3. In § 4 the definition of degree of mapping will be extended to any continuous mapping of an open set G in E^m. In § 5 a product theorem of degree of mapping, concerning the composition of two mappings, will be proved.

2. Auxiliary theorems.

THEOREM 1.[1] *Let* $x'=f(x)$ *be a regular mapping of an open set* G *in* E^m. *Then the crease of* f *is a set of (m-dimensional Lebesgue) measure zero in* E^m.

Proof. We divide G into an enumerable set of closed cubes Q: $|x_i-a_i|\leqq l$ $(i=1,2,\cdots,m)$. Then it is sufficient to prove that the measure of the crease of f on Q is zero. For any $\epsilon>0$, there exists a natural number n such that Q can be divided into n^m equal small closed cubes Q_ν $(\nu=1,\cdots,n^m)$ in each of which we have

$$f_i(x)=f_i(\alpha)+\sum_{j=1}^{m}\partial_jf_i(\alpha)(x_j-\alpha_j)+\eta_i(x),$$

where $\sum_{i=1}^{m}\eta^2<(\epsilon/n)^2$ and α is an arbitrary point in Q_ν. Let $\hat{Q}_\nu$ be the image of Q_ν under the linear transformation $x'_i=f_i(\alpha)+\sum_{j=1}^{m}\partial_jf_i(\alpha)(x_j-\alpha_j)$. The diameter of $\hat{Q}_\nu$ is not greater than $2Ll/n$, where $L=\mathrm{Max}_Q(\sum_{i,j=1}^{m}\{\partial_jf_i(x)\}^2)^{1/2}$. If Q_ν contains a critical point, we take it as α. Then Q_ν is contained in an

---

[1] This is the simplest special case of a theorem of A. Sard, *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 883-890.

$(m-1)$-dimensional linear manifold, and $f(Q_\nu)$ lies in the ϵ/n-neighborhood of $\hat{Q}$. Therefore the measure of $f(Q_\nu)$ is smaller than $\{2n^{-1}(Ll+\epsilon)\}^{m-1}2\epsilon n^{-1}$. Consequently the measure of the crease in $f(Q)$ is smaller than $2^m(Ll+\epsilon)^{m-1}\epsilon$, and since ϵ is arbitrary small, it must be zero.

As an application of Theorem 1 we obtain immediately:

THEOREM 2. *Let M be any set of points in E^m which can be represented by the equations*

$$x_\mu = \phi_\mu(s_1, \cdots, s_l) \qquad (\mu = 1, \cdots, m), \quad 0 < l < m,$$

where $\phi_\mu(s_1, \cdots, s_l)$ are continuously differentiable in an open domain of an $l(< m)$-dimensional Euclidean space. Then M is a set of (m-dimensional Lebesgue) measure zero in E^m.

THEOREM 3. *Let G be a bounded open set in E^m and $x' = F(x)$ be a regular mapping of G into E^m. Then for any point $a \in E^m$, neither on $F(\bar{G} - G)$ nor on the crease of F, there exists a positive number ϵ with the following property:*

For any regular mapping $x' = f(x)$ of G into E^m such that $|f(x) - F(x)| < \epsilon^2$ for $x \in \bar{G}$ and $|\partial_j f(x) - \partial_j F(x)| < \epsilon$ $(j = 1, \cdots, m)$ for $x \in G_\epsilon$, where $G_\epsilon = G \cap \{x \mid \mathrm{dist}(x, \bar{G} - G) > \epsilon\}$, the equality $A[a, G, f] = A[a, G, F]$ holds.

Proof. There are only a finite number of points x of G such that $F(x) = a$. Let $p^1, \cdots, p^N$ be all such points. Let K be the set of critical points of F, and put

$$\Delta = \mathrm{Min}\{|p^\mu - p^\nu| \ (\mu \neq \nu), \quad \mathrm{dist}(p^\mu, K \cup (\bar{G} - G))\};$$

then $\Delta > 0$. Let α be a positive number such that $\alpha < \Delta/2$, and $U_\alpha(p^\nu)$ be the α-neighborhood of p^ν, and put

$$M_\alpha = \mathrm{Min}\{|F(x) - a| \mid x \in \bar{G} - \bigcup_{\nu=1}^{N} U_\alpha(p^\nu)\};$$

then $U_\alpha(p^\nu) \subset G_\alpha$, where $G_\alpha = G \cap \{x \mid \mathrm{dist}(x, \bar{G} - G) > \alpha\}$, and $M_\alpha > 0$.

Since $\det(\partial_x F(x)) \neq 0$ for $x \in \bar{U}_\alpha(p^\nu)$, there exists an $\epsilon' > 0$ such that $\det(\partial_x f(x)) \neq 0$ holds for any regular mapping f of $U_\alpha(p^\nu)$ into E^m provided that $|\partial_j f(x) - \partial_j F(x)| < \epsilon'$ $(j = 1, \cdots, m)$.

Now let f be any regular mapping of G into E^m such that $|f(x) - F(x)|$

---

[2] By $|p - q|$ we denote the distance $(\sum_{i=1}^{m} (p_i - q_i)^2)^{1/2}$ of two points p and q in E^m.

$< M_\alpha$ for $x \varepsilon \bar{G}$ and $|\partial_j f(x) - \partial_j F(x)| < \epsilon'$ $(j = 1, \cdots, m)$ for $x \varepsilon G_\alpha$. We put $f_t(x) = (1 - t)F(x) + tf(x)$. Then we have

$$\det(\partial_x f_t(x)) \neq 0 \text{ for } 0 \leqq t \leqq 1, x \varepsilon U_\alpha(p^\nu),$$

and all roots of $f_t(x) = a$ lie in $\bigcup_{\nu=1}^{N} U_\alpha(p^\nu)$.

Let τ be any value of t from $0 \leqq t \leqq 1$ and p_0 be a point of G such that $f_\tau(p_0) = a$. Then p_0 belongs to a $U_\alpha(p^\nu)$ and by the theory of implicit functions there exists a neighborhood of $t = \tau$ such that the equation $f_t(x) = a$ has exactly one solution $x = p(t)$ in a sufficiently small neighborhood of p_0. This solution $p(t)$ depends on t continuously and can be prolongated for the whole interval $0 \leqq t \leqq 1$. In fact, let $p(t)$ be continuous and $p(t) \varepsilon U_\alpha(p^\nu)$ for $\tau \leqq t < T$, and p^* be any accumulation point of $p(t)$ for $t \to T$. Then $p^* \varepsilon \bar{U}_\alpha(p^\nu)$ and $f_T(p^*) = a$. Since $\det(\partial_x f_T(p^*)) \neq 0$, then by the theory of implicit functions there exists a unique solution $x = p^*(t)$ of $f_t(x) = a$ in a sufficiently small neighborhood of p^* for t sufficiently near to T. As $p^*(t)$ is continuous and the solution of $f_t(x) = a$ is unique in a small neighborhood of p^*, it must coincide with $p(t)$ for $t < T$ sufficiently near to T. Thus $p(t)$ is continuous for $\tau \leqq t \leqq T$ and can be prolongated beyond T if $T < 1$. This and a similar consideration for $t \leqq \tau$ show that $p(t)$ can be prolongated for $0 \leqq t \leqq 1$, where it is continuous. Since the solutions of $f_t(x) = a$ are isolated and continuous for $0 \leqq t \leqq 1$, the number of the solutions of $f_t(x) = a$ is constant in each $U_\alpha(p^\nu)$ for $0 \leqq t \leqq 1$. Hence $f_t(x) = a$ has exactly one solution in each $U_\alpha(p^\nu)$ and $\det(\partial_x f_t)$ keeps the same sign in each $U_\alpha(p^\nu)$ for $0 \leqq t \leqq 1$. Then $\Delta[a, G, f_t]$ is constant for $0 \leqq t \leqq 1$. The proof is thus established, if we take $\epsilon = \text{Min}\{\alpha, M_\alpha, \epsilon'\}$.

THEOREM 4. *Let G be a bounded open set in E^m, $x' = F(x)$ a regular mapping of G into E^m, $a(t) \varepsilon E^m$ continuously differentiable in $0 \leqq t \leqq 1$, and $a(t) \bar{\varepsilon} F(\bar{G} - G)$. Then there exists for any $\epsilon > 0$ such that $\epsilon < \text{dist}(a(t), F(\bar{G} - G))$ a regular mapping $x' = f(x)$ of G into E^m with the following properties:*

i) *$|f(x) - F(x)| < \epsilon$ for $x \varepsilon \bar{G}$ and $|\partial_j f(x) - \partial_j F(x)| < \epsilon$ $(j = 1, \cdots, m)$ for $x \varepsilon G_\epsilon$.*[3]

ii) *For every x such that $f(x) = a(t)$ $(0 \leqq t \leqq 1)$, there holds the inequality*

$$\text{Rank of } (\partial_x f(x)) \geqq m - 1, \textit{ where } (\partial_x f) = (\partial_j f_i)_{i,j=1,\ldots,m}.$$

---

[3] $G_\epsilon = G \cap \{x \mid \text{dist}(x, \bar{G} - G) > \epsilon\}$.

iii) *For any fixed t the set of points x such that $f(x) = a(t)$ has no accumulation point in G.*

Proof. We can find a regular mapping $x' = \phi(x)$ of G into E^m, such that $\phi(x)$ is 3 times continuously differentiable in G, $|\phi(x) - F(x)| < \epsilon/2$ for $x \varepsilon \bar{G}$ and $|\partial_j\phi(x) - \partial_j F(x)| < \epsilon/2$ for $x \varepsilon G_\epsilon$. We put

$$f_i(x) = \phi_i(x) + \alpha_i + \sum_{\nu=1}^{m} (\beta_{i,\nu}x_\nu + 2^{-1}\gamma_{i,\nu}x^2_\nu). \tag{1}$$

Then we can find $\delta > 0$, such that $f(x)$ defined by (1) satisfies the statement i) if the conditions $|\alpha_i| < \delta$, $|\beta_{i,\nu}| < \delta$, $|\gamma_{i,\nu}| < \delta$ are satisfied.

Let us prove that the set of (α, β, γ) such that $f(x)$ defined by (1) does not satisfy ii) has the measure zero in $E^q{}_{(\alpha,\beta,\gamma)}$ $(q = m + 2m^2)$. If ii) does not hold, then there exists a point x and a value of t for which $f(x) = a(t)$ and certain $l \geqq 2$ columns of the matrix $(\partial_x f)$ are linear combinations of the other $m - l$ columns. Let us suppose that these are the last l columns. Then there exist $l(m-l)$ numbers $\lambda_{\mu,\nu}$ $(\mu = 1, \cdots, l,$ $\nu = 1, \cdots, m-l)$ such that

$$\partial_{m-l+\mu}f_i = \sum_{\nu=1}^{m-l} \lambda_{\mu,\nu}\partial_\nu f_i \qquad (i = 1, \cdots, m).$$

Then by (1):

$$\beta_{i,m-l+\mu} = \sum_{\nu=1}^{m-l} \lambda_{\mu,\nu}(\beta_{i,\nu} + \partial_\nu\phi_i(x) + \gamma_{i,\nu}x_\nu)$$
$$- \partial_{m-l+\mu}\phi_i(x) - \gamma_{i,m-l+\mu}x_{m-l+\mu},$$

and

$$\alpha_i = a_i(t) - \phi_i(x) - \sum_{\nu=1}^{m} (\beta_{i,\nu}x_\nu + 2^{-1}\gamma_{i,\nu}x_\nu^2).$$

This shows that (α, β, γ) can be represented by continuously differentiable functions of $m + 2m^2 - l^2 + 1 < q$ arguments $t, x_j, \lambda_{\mu,\nu}, \beta_{i,\nu}, \gamma_{i,\nu}$ $(i, j = 1, \cdots, m, \mu = 1, \cdots, l, \nu = 1, \cdots, m-l)$. Thus by Theorem 2 the set of points (α, β, γ) such that $f(x)$ does not satisfy ii) has the measure zero in $E^q{}_{(\alpha,\beta,\gamma)}$, since there are only a finite number of ways to choose $l \geqq 2$ columns from $(\partial_x f)$.

Next we shall prove that the set of (α, β, γ) for which the statement ii) holds, but not iii), has also the measure zero in $E^q{}_{(\alpha,\beta,\gamma)}$. Suppose that this occurs for $t = t^*$ and put $a(t^*) = a$. Then the equation $f(x) = a$ has an infinite number of roots $x = p^n = (p_1{}^n, \cdots, p_m{}^n)$, which converges to a point $p = (p_1, \cdots, p_m) \varepsilon G$. By ii) there exist $m - 1$ components of $f(x)$, say $f_1, \cdots, f_{k-1}, f_{k+1}, \cdots, f_m$, such that $D(f_1, \cdots, f_{k-1}, f_{k+1}, \cdots, f_m/x_1,$

$\cdots, x_{m-1}) \neq 0$ for $x = p$. If we put $y_i = f_i(x)$ for $i \neq k$ and $y_k = x_m$, then $D(y/x)_{x=p} = \pm D(f_1, \cdots, f_{k-1}, f_{k+1}, \cdots, f_m / x_1, \cdots, x_{m-1}) \neq 0$, and by a theorem of implicit functions, a neighborhood of p in $E_{(x)}{}^m$ corresponds homeomorphically to a neighborhood of $(a_1, \cdots, a_{k-1}, p_m, a_{k+1}, \cdots, a_m)$ in $E_{(y)}{}^m$. Let $x = \psi(y)$ be the inverse mapping (3 times continuously differentiable) and put $p_m{}^n = s_n$, $p_m = \sigma$, $\psi_\mu(a_1, \cdots, a_{k-1}, s, a_{k+1}, \cdots, a_m) = h_\mu(s)$ $(\mu = 1, \cdots, m)$. Then we have $\lim s_n = \sigma$, $p_\mu{}^n = h_\mu(s_n)$, $p_\mu = h_\mu(\sigma)$ and $f_i(h(s_n)) = f_i(h(\sigma)) = a_i$. Thus, if we put $h'_\mu(\sigma) = \lambda_\mu$ and $h''_\mu(\sigma) = \lambda'_\mu$, we have

$$\frac{d}{ds} f_i(h(s))_{s=\sigma} = \sum_{\nu=1}^{m} \lambda_\nu (\partial_\nu f_i)_{x=p} = 0, \tag{2}$$

and

$$\frac{d^2}{ds^2} f_i(h(s))_{s=\sigma} = \sum_{\mu,\nu=1}^{m} \lambda_\mu \lambda_\nu (\partial^2{}_{\mu,\nu} f_i)_{x=p} + \sum_{\mu=1}^{m} \lambda'_\mu (\partial_\mu f_i)_{x=p} = 0. \tag{3}$$

But by definition we have $h_m(s) = s$. Thus we get $\lambda_m = h'_m(\sigma) = 1$ and $\lambda'_m = h''_m(\sigma) = 0$. Hence from (1) and (3) we obtain

$$\begin{aligned} \gamma_{i,m} = -\sum_{\mu=1}^{m-1} \{ \sum_{\nu=1}^{m-1} \lambda_\mu \lambda_\nu \partial^2{}_{\mu,\nu} \phi_i(p) + 2\lambda_\mu \partial^2{}_{\mu,m} \phi_i(p) \\ + \lambda_\mu \gamma_{i,\mu} + \lambda'_\mu (\partial_\mu \phi_i(p) + \beta_{i,\mu} + \gamma_{i,\mu} p_\mu) \} - \partial^2{}_{m,m} \phi_i(p). \end{aligned}$$

And by (1) and (2) we get

$$\beta_{i,m} = -\sum_{\mu=1}^{m-1} \lambda_\mu (\partial_\mu \phi_i(p) + \beta_{i,\mu} + \gamma_{i,\mu} p_\mu) - \gamma_{i,m} p_m - \partial_m \phi_i(p).$$

Thus (α, β, γ) can be represented by continuously differentiable functions of $m + 2m^2 - 1 = q - 1$ arguments t, p_j, λ_μ, λ'_μ, $\beta_{i,\mu}$, $\gamma_{i,\mu}$ $(i, j = 1, \cdots, m,$ $\mu = 1, \cdots, m - 1)$. Therefore by Theorem 2 the set of points (α, β, γ) such that ii) holds but not iii) has measure zero in $E^q{}_{(\alpha,\beta,\gamma)}$.

In conclusion we have that, under the conditions $|\alpha_i| < \delta$, $|\beta_{i,\mu}| < \delta$, $|\gamma_{i,\mu}| < \delta$, except for a set of points (α, β, γ) of measure zero the statements ii) and iii) are satisfied by $f(x)$ defined by (1).

3. Proof of the continuity of the degree of mapping.

THEOREM 5. *Let $x' = F_t(x)$ be a regular mapping of a bounded open set G in E^m into E^m and $F_t(x)$ be a continuous function of (t, x) for $0 \leqq t \leqq 1$, $x \,\varepsilon\, \bar{G}$. If $a(t)\, (\varepsilon E^m)$ is continuous and $a(t) \,\bar{\varepsilon}\, F_t(\bar{G} - G)$ for $0 \leqq t \leqq 1$, then $A[a(t), G, F_t]$ is constant for $0 \leqq t \leqq 1$ (except for those values of t such that $a(t)$ is on the crease of F_t).*

Proof. For the case $m = 1$ the proof is not difficult, because one can approximate the mapping function including its derivative by a function which has only a finite number of critical points. Therefore we assume that the theorem holds for $m = 1$, and we suppose that $m > 1$.

At first let us consider the case that $F_t(x)$ does not depend on t, and $a(t)$ is continuously differentiable. It is sufficient to prove $A[a(0), G, F] = A[a(1), G, F]$. Let ϵ_0 be the distance between the curve $x = a(t)$ and $F(\bar{G} - G)$. Then by Theorem 3 and Theorem 4 there exists a regular mapping $x' = f(x)$ of G into E^m with the following properties:

i) $| f(x) - F(x) | < \epsilon_0$ for $x \varepsilon \bar{G}$.

ii) $A[a(i), G, f] = A[a(i), G, F] \quad (i = 0, 1)$.

iii) Rank of $(\partial_x f) \geqq m - 1$ for any x such that $f(x) = a(t)$ $(0 \leqq t \leqq 1)$.

iv) For any fixed t in $0 \leqq t \leqq 1$ the set of the roots of $f(x) = a(t)$ has no accumulation point in G.

Now we take any fixed value of t in $0 \leqq t \leqq 1$ and put $a(t) = a$. Then from i) and the definition of ϵ_0 we have $a \bar{\varepsilon} f(\bar{G} - G)$. And from iv) the equation $f(x) = a$ has only a finite number of roots $c^1, \cdots, c^k$. We can find a positive number Δ such that, if U_ν is any neighborhood of c^ν with diameter smaller than Δ, the U_ν all lie in G and have no common points with one another. Since $a \bar{\varepsilon} f(\bar{G} - U_1 - \cdots - U_k)$, all roots of $f(x) = a'$ lie in $U_1 \cup \cdots \cup U_k$ for any point a' with

$$| a - a' | < \operatorname{dist}[a, f(\bar{G} - U_1 - \cdots - U_k)] = \Delta', \tag{1}$$

and

$$A[a', G, f] = \sum_{\nu=1}^{k} A[a', U_\nu, f], \tag{2}$$

if a' is not on the crease of f.

Let us consider each $A[a', U_\nu, f]$ separately. From iii) there exist $m - 1$ components of f, say $f_1, \cdots, f_{l-1}, f_{l+1}, \cdots, f_m$, such that

$$D(f_1, \cdots, f_{l-1}, f_{l+1}, \cdots, f_m / x_1, \cdots, x_{m-1}) \neq 0 \text{ at } x = c^\nu.$$

If we put

$$y_i = f_i(x) \text{ for } i \neq l \text{ and } y_l = x_m, \tag{3}$$

then $D(y/x) \neq 0$ at $x = c^\nu$. By a theorem of implicit functions we can find a positive number δ, such that the neighborhood V_ν of the point $(a_1, \cdots, a_{l-1}, c_m{}^\nu, a_{l+1}, \cdots, a_m)$ in $E_{(y)}{}^m$ defined by

$$V_\nu \colon \alpha_i < y_i < \beta_i \qquad (i = 1, \cdots, m),$$

where $\alpha_i = a_i - \delta$, $\beta_i = a_i + \delta$ for $i \neq l$ and $\alpha_l = c_m{}^\nu - \delta$, $\beta_l = c_m{}^\nu + \delta$, corresponds homeomorphically to a neighborhood U_ν of c^ν in $E_{(x)}{}^m$ whose diameter is smaller than Δ. Let us denote its inverse mapping by $x = \phi(y)$, and put $f(\phi(y)) = g(y)$. We can suppose that the correspondence is continuous on the boundary $\bar{U}_\nu - U_\nu = \phi(\bar{V}_\nu - V_\nu)$ and that $D(\phi/y) \neq 0$ in V_ν. If we put $\operatorname{sign}\{D(\phi/y)\} = e_\nu$, which is constant in V_ν, we have by the definition of the degree of mapping

$$(4) \qquad A[a', U_\nu, f] = e_\nu A[a', V_\nu, g].$$

From $x' = f(x) = g(y)$ and (3) we have $g_i(y) = y_i$ for $i \neq l$. Therefore $D(g/y) = \partial_l g_l(y)$. Consequently we get

$$(5) \qquad A[a', V_\nu, g] = A[a'_l, (\alpha_l, \beta_l), g_{l,(a')}],$$

where the right-hand side means the degree of mapping by the one-dimensional transformation $x'_l = g_l(a'_1, \cdots, a'_{l-1}, y_l, a'_{l+1}, \cdots, a'_m) = g_{l,(a')}(y_l)$. Let a'' be any point in $|a'' - a| < \Delta'$ not on the crease of f; then the segment $a(s) = (1-s)a' + sa''$ $(0 \leqq s \leqq 1)$ lies in U_ν and $a(s) \bar{\varepsilon} g(\bar{V}_\nu - V_\nu)$. Therefore by the hypothesis $A[a_l(s), (\alpha_l, \beta_l), g_{l,[s]}]$ is constant for $0 \leqq s \leqq 1$, where $g_{l,[s]}$ means the one-dimensional transformation $x'_l = g_{l(a(s))}(y_l)$. Thus by (5) we have

$$(6) \qquad A[a', V_\nu, g] = A[a'', V_\nu, g].$$

From (2), (4), and (6) we get

$$A[a', G, f] = A[a'', G, f].$$

This shows that $A[a', G, f]$ does not depend on a', when a' is sufficiently near a point of the curve $x = a(t)$ $(0 \leqq t \leqq 1)$. By using Borel's covering theorem for the curve $x = a(t)$ we obtain $A[a(0), G, f] = A[a(1), G, f]$. And by ii) we get $A[a(0), G, F] = A[a(1), G, F]$. This completes the first part of the proof.

Now we shall consider the case that $F_t(x)$ depends on t. Let τ be any value of t in $0 \leqq t \leqq 1$ and put $\operatorname{dist}[a(\tau), F_\tau(\bar{G} - G)] = \Delta$. Then there exists a $\delta > 0$, such that for $|t - \tau| < \delta$, $x \in \bar{G}$, the inequalities

$$|F_t(x) - F_\tau(x)| < \Delta/2 \quad \text{and} \quad |a(t) - a(\tau)| < \Delta/2$$

hold. Let t_0, t_1 be any two values of t in $|t - \tau| < \delta$ and put

$$a[s] = (1-s)a(t_0) + sa(t_1), \qquad F_{(s)}(x) = (1-s)F_{t_0}(x) + sF_{t_1}(x).$$

Then we have $a[s] \bar{\varepsilon} F_{(s)}(\bar{G} - G)$ for $0 \leqq s \leqq 1$. Let H be the open set in $E_{(s,x)}{}^{m+1}$ defined by $-\epsilon < s < 1 + \epsilon$ $(\epsilon > 0)$, $x \in G$, and $(s', x') = \Phi(s, x)$

be the transformation in $E_{(s,x)}{}^{m+1}$ defined by $s' = s$, $x' = F_{(s)}(x)$. Then Φ is a regular mapping of H into $E_{(s,x)}{}^{m+1}$, and $a^*(t) = (t, a[t])$ does not touch $\Phi(\bar{H} - H)$ for $0 \leqq t \leqq 1$. Therefore $A^{m+1}[a^*(t), H, \Phi]$ is constant for $0 \leqq t \leqq 1$ by the first part of the proof. The superscript of A denotes the dimension of the space in which the mapping is considered. But one can easily prove also that

$$A^{m+1}[a^*(t), H, \Phi] = A^m[a[t], G, F_{(t)}].$$

Thus we get $A[a(t_0), G, F_{t_0}] = A[a(t_1), G, F_{t_1}]$. As t_0 and t_1 are arbitrary values in $|t - \tau| < \delta(\tau)$ and τ in $0 \leqq t \leqq 1$, then by the covering theorem of Borel we obtain the constancy of $A[a(t), G, F_t]$ for $0 \leqq t \leqq 1$. Thus the proof is complete.

Remark. Let $x' = F(x)$ be a regular mapping of a bounded open set G in E^m and a_1 and a_2 be any two points such that $|a_i - a| < \operatorname{dist}[a, F(\bar{G} - G)]$. Then $A[a_1, G, F] = A[a_2, G, F]$, if a_1 and a_2 are not on the crease of F. From this we can define $A[a, G, F]$, *even when a is on the crease of F, only provided a is not in $F(\bar{G} - G)$.* For, in any small neighborhood V of a there is a point a' not on the crease of F; and $A[a', G, F]$ is independent of a', when a' is in a sufficiently small neighborhood V of a. Therefore *Theorem 5 holds even when $a(t)$ is on the crease of F_t.*

4. The degree of mapping of continuous transformations.

Now we shall extend the definition of degree of mapping to any continuous mapping of an open set in E^m. First let us consider a bounded open set and a mapping which is continuous on its closure. *Let G be a bounded open set in E^m, $x' = f(x)$ a continuous mapping of $\bar{G}$ into E^m and a a point not on $f(\bar{G} - G)$. Then $A[a, G, F]$ has the same value for any regular mapping $x' = F(x)$ of G into E^m, if F satisfies the condition*

$$(1) \qquad |F(x) - f(x)| < \operatorname{dist}[a, f(\bar{G} - G)] \qquad \textit{for } x \in \bar{G}.$$

Thus we define the degree of mapping of G at a (by f) by

$$A[a, G, f] = A[a, G, F].$$

In order to prove that $A[a, G, F]$ has the same value for any regular mapping F of G under the conidtion (1), let us take two regular mappings F_0 and F_1 which satisfy the condition (1), and put $F_t(x) = (1 - t)F_0(x) + tF_1(x)$. Then $F_t(x)$ satisfies the condition (1) for $0 \leqq t \leqq 1$, and hence $a \bar{\varepsilon} F_t(\bar{G} - G)$. Thus, by Theorem 5, $A[a, G, F_t]$ is constant for $0 \leqq t \leqq 1$, and we obtain $A[a, G, F_0] = A[a, G, F_1]$.

The degree of mapping thus defined also satisfies the properties i), ii), iii), and iv) in § 1, as one can prove easily, when $f(x)$ is continuous in $\bar{G}$.

Next let us consider a more general case: Let G be an open set in E^m and $x' = f(x)$ be a continuous mapping of G into E^m. *Let* $(f; G)$ *be the set of all accumulation points of* $\{f(p_n)\}$ *where* $\{p_n\}$ *is any sequence of points of* G *having no accumulation point in* G. If G is bounded and f is continuous on $\bar{G}$ (the closure of G), then $(f; G) = f(\bar{G} - G)$. If $a \bar{\varepsilon} (f; G)$, then the set R of all roots of the equation $f(x) = a$ in G is a bounded closed set in E^m. Let G_i $(i = 1, 2)$ be any two bounded open sets with the properties $R \subset G_i$ and $\bar{G}_i \subset G$. Then f is continuous on $\bar{G}_i$ and it is easy to see that

$$A[a, G_1, f] = A[a, G_2, f].$$

Thus *we define* $A[a, G, f]$ *by* $A[a, G, f] = A[a, G_1, f]$ *where* G_1 *is any bounded open set with the properties* $R \subset G_1$ *and* $\bar{G}_1 \subset G$, *where* R *is the set of all roots of* $f(x) = a$ *in* G.

The properties i), ii), and iii) in § 1 hold also for the degree of mapping of this kind, if one replaces $f(\bar{G} - G)$ by $(f; G)$. To extend the statement iv) to this case, we define $(f_t; G, \tau)$ as follows:

Let f_t *be a continuous mapping of* G *into* E^m *depending on the parameter* t, $0 \leqq t \leqq 1$. *Then* $(f_t; G, \tau)$ *means the set of all accumulation points of* $\{f_{t_n}(p_n)\}$ *where* $\lim t_n = \tau$ *and* $\{p_n\}$ *is any sequence of points of* G *having no accumulation point in* G. If G is bounded and f_t is a continuous function of (t, x) for $0 \leqq t \leqq 1$, $x \varepsilon \bar{G}$, then we have $(f_t; G, \tau) = f_\tau(\bar{G} - G)$ for any $0 \leqq \tau \leqq 1$.

To prove the statement iv) where $f_t(\bar{G} - G)$ is replaced by $(f_t; G, \tau)$, we use the following lemma which one can prove easily:

If $a \bar{\varepsilon} (f_t; G, \tau)$, *then there exists a neighborhood* V *of* a, *a positive number* δ *and a bounded open set* G_1, *such that* $\bar{G}_1 \subset G$ *and* $V \cap f_t(G - G_1)$ *is empty for* $| t - \tau | < \delta$.

We can take δ so small that $a(t) \varepsilon V$ also holds for $| t - \tau | < \delta$. Thus we have for $| t - \tau | < \delta$,

$$A[a(t), G, f_t] = A[a(t), G_1, f_t].$$

As $\bar{G}_1 - G_1 \subset G - G_1$, we have also $a(t) \bar{\varepsilon} f_t(\bar{G}_1 - G_1)$ for $| t - \tau | < \delta$. Then $A[a(t), G_1, f_t]$ is constant for $| t - \tau | < \delta$. Consequently $A[a(t), G, f_t]$ is constant for $| t - \tau | < \delta$. By Borel's covering theorem applied to the closed interval $0 \leqq t \leqq 1$ *we obtain statement* iv), *replacing* $a(t) \bar{\varepsilon} f_t(\bar{G} - G)$ *by* $a(\tau) \bar{\varepsilon} (f_t; G, \tau)$.

5. The product theorem.

THEOREM 6. *Let G be an open set in E^m and $x' = f(x)$ a continuous mapping of G into E^m. Let H be an open set containing $f(G)$ and H_i $(i = 1, 2, \cdots)$ the components of the open set $H - (f; G)$. Let $x' = \phi(x)$ be a continuous mapping of H into E^m and a a point of E^m outside of $(\phi f; G) \cup (\phi; H)$. Then we have*

$$A[a, G, \phi f] = \sum_i A[a, H_i, \phi] A[b_i, G, f], \tag{1}$$

where b_i is an arbitrary point of H_i.

Proof. $A[p, G, f]$ is constant when p varies on a domain H_i. Let D_k be the sum of those domains H_i such that $A[p, G, f] = k$ for $p \varepsilon H_i$. Then D_k is the set of all points $p \varepsilon H$ such that $A[p, G, f] = k$.

Now let R be the set of all roots of $\phi(x) = a$. Then since R is compact and $R \cap (f; G)$ is empty (because $a \bar{\varepsilon} \phi(H \cap (f; G)) \subset (\phi f; G)$), R can be covered by a finite number of H_i. Thus we get by property iii),

$$\sum_i A[a, H_i, \phi] A[b_i, G, f] = \sum_k A[a, D_k, \phi] k. \tag{2}$$

Therefore it suffices to prove the relation

$$A[a, G, \phi f] = \sum_k A[a, D_k, \phi] k. \tag{3}$$

The relation (3) is true, if G and H are bounded, f and ϕ are regular mappings, and a is not on the crease of ϕf, as one can prove easily from the definition of the degree of mapping.

In order to discuss the general case we first assume that G and H are bounded, that f and ϕ are continuous on $\bar{G}$ and $\bar{H}$ respectively and that $f(\bar{G}) \subset H$. Let us take regular transformations f^* and ϕ^* sufficiently near to f and ϕ respectively such that a is not on the crease of $\phi^* f^*$:

$$| f^*(x) - f(x) | < \operatorname{dist}[R, f(\bar{G} - G)] \text{ for } x \varepsilon \bar{G},$$

$$| \phi^*(x) - \phi(x) | < \operatorname{dist}[a, \phi f^*(\bar{G} - G) \cup \phi(\bar{H} - H)] \text{ for } x \varepsilon \bar{H},$$

and

$$| \phi^* f^*(x) - \phi f(x) | < \operatorname{dist}[a, \phi f(\bar{G} - G)] \text{ for } x \varepsilon \bar{G}.$$

Thus, if we denote by D^*_k the set of all points $p \varepsilon H$ such that $A[p, G, f^*] = k$, we have

$$\operatorname{dist}[a, \phi f^*(\bar{G} - G) \cup \phi(\bar{H} - H)] \leqq \operatorname{dist}[a, \phi(\bar{D}^*_k - D^*_k)];$$

$$A[a, D^*_k, \phi^*] = A[a, D^*_k, \phi] = A[a, D_k, \phi], \tag{4}$$

since $D^*_k \cap R = D_k \cap R$ (because $A[p, G, f^*] = A[p, G, f]$ for $p \varepsilon R$); and

$$A[a, G, \phi^* f^*] = A[a, G, \phi f]. \tag{5}$$

But, as we have already said, we have

$$A[a, G, \phi^* f^*] = \sum_k [a, D^*_k, \phi^*]k. \tag{6}$$

Thus from (4), (5), and (6) we get

$$A[a, G, \phi f] = \sum_k A[a, D_k, \phi]k. \tag{7}$$

Next we discuss the general case, when G and H are not necessarily bounded and f and ϕ are continuous respectively on G and H. Let $\{G_n\}$ and $\{H_n\}$ be sequences of bounded open sets such that $\bar{G}_n \subset G_{n+1}$, $\lim G_n = G$, $f(\bar{G}_n) \subset H_n$, $\bar{H}_n \subset H_{n+1}$, $\lim H_n = H$, and $D_k{}^n$ the set of all points $p \varepsilon H_n$ such that $A[p, G_n, f] = k$. Then $f(\bar{G}_n - G_n)$ approaches $(f; G)$ as $n \rightarrow \infty$, and $\underline{\lim}\, D_k{}^n \supseteq D_k$ (as point sets). Since R can be covered by a finite number of D_k, for example by $D_{-l}, \cdots, D_l$, and each $D_k \cap R$ is compact, it follows that R can be covered by $D_{-l}{}^n, \cdots, D_l{}^n$ for a sufficiently large n. Thus we have

$$A[a, G_n, \phi f] = A[a, G, \phi f] \text{ and } A[a, D_k{}^n, \phi] = A[a, D_k, \phi]$$

for a certain fixed n and all k. Therefore (7) holds for the general case. Thus from (2) and (7) we obtain (1).

OSAKA UNIVERSITY,
OSAKA, JAPAN.

DEGREE OF MAPPING IN CONVEX LINEAR TOPOLOGICAL SPACES.*

By Mitio Nagumo.

J. Leray and J. Schauder have developed a theory of the degree of mapping for a completely continuous movement[1] in a Banach space (1934) [1]. Before this, Schauder gave a theorem on the invariance of domain under a one-one completely continuous movement in a weakly compact Banach space (1929) [6], (1932) [7]. These works contain important applications to the theory of partial differential equations of elliptic type. Leray also gave a product theorem concerning the degree of mapping for the composition of completely continuous movements in general Banach spaces and extended the theorem of Schauder on the invariance of domain to the case of general Banach spaces (1935) [2]. However, the treatments [1] and [2] seem not to be complete in detail. In this paper, following Leray and Schauder, I wish to establish a theory of the degree of mapping in the case of convex linear topological spaces and prove a theorem on the invariance of domain in the case of convex linear metric complete spaces.[2]

I am much obliged to the referee of this Journal for several corrections of my manuscript and for references to the literature.

1. Preliminary notions.

1. 1. First we shall explain the notion of the degree of mapping in a finite dimensional Euclidean space. Let G be an open set in an m-dimensional Euclidean space E^m and $x' = f(x) = \{f_i(x) \mid i = 1, \cdots, m\}$ $(x = (x_1, \cdots, x_m))$ be a continuous mapping of $\bar{G}$ (the closure of G) into E^m such that $f(x) - x$ is bounded on $\bar{G}$. Let a be a point of E^m *not in* $f(\bar{G} - G)$ (the image of $\bar{G} - G$ by f). Then there will be determined an integer $A[a, G, f]$, called the *degree of mapping* of G at a by f, with the following properties:

i) *If f is the identical mapping $f(x) = x$, then*

$A[a, G, f] = 1$ *when* $a \varepsilon G$, $A[a, G, f] = 0$ *when* $a \bar{\varepsilon} \bar{G}$.[3]

---

* Received March 6, 1950; revised September 4, 1950.

[1] I. e., a transformation of the form $x' = x + f(x)$, where f is completely continuous.

[2] I am also informed about the work of E. Rothe on the same subject [5], but I must sorrily confess that I have had no chance to see it.

[3] $\bar{\varepsilon}$ means the negation of ε.

ii) *If* $A[a, G, f] \neq 0$, *then there exists a point* $x \varepsilon G$ *such that* $f(x) = a$.

iii) *If* G *is divided into open sets* $G_1, \cdots, G_k$, *i. e.* $G \supset \bigcup_{i=1}^{k} G_i$, $\bar{G} = \bigcup_{i=1}^{k} \bar{G}_i$, *and* $G_i \cap G_j = 0$ *(empty set)* $(i \neq j)$, *and if* $a \bar{\varepsilon} f(\bar{G}_i - G_i)$ *for any* i, *then*

$$A[a, G, f] = \sum_{i=1}^{k} A[a, G_i, f].$$

iv) *If* $f_t(x) - x$ *is a bounded continuous function of* (t, x) *for* $0 \leqq t \leqq 1$, $x \varepsilon \bar{G}$, *if* $a(t) (\varepsilon E^m)$ *is continuous, and if* $a(t) \bar{\varepsilon} f_t(\bar{G} - G)$ *for all* t *in* $0 \leqq t \leqq 1$, *then* $A[a(t), G, f_t]$ *is constant for* $0 \leqq t \leqq 1$.

v) *Let* a *be a point not on* $f(\bar{G} - G)$, X *the set of all roots of the equation* $f(x) = a$ *in* G, *and* G_0 *any open set such that* $X \subset G_0 \subset G$, *then*

$$A[a, G_0, f] = A[a, G, f].$$

Remark. v) follows from ii) and iii).

The existence and uniqueness of $A[a, G, f]$ satisfying the above conditions can be verified, if we use simplicial mappings for approximations of f (at first for bounded G and then for general G). But I may refer to [3] in which the existence of $A[a, G, f]$ is given, based on infinitesimal analysis but free from the notion of simplicial mapping. (It will be easily seen that $f(\bar{G} - G) = (f; G)$ and $f_\tau(\bar{G} - G) = (f_t; G, \tau)$ in the notation of [3].)

From iv) we have

COROLLARY. *If* f_0 *and* f_1 *are continuous transformations of* $\bar{G}$ *into* E^m *such that* $|f_1(x) - f_0(x)| < \operatorname{dist}(a, f_0(\bar{G} - G))$,[4] *then*

$$A[a, G, f_1] = A[a, G, f_0].$$

Proof. Put $f_t(x) = (1 - t) f_0(x) + t f_1(x)$ and apply iv).

1.2. A linear topological space is a linear set on which is imposed a topology in such a fashion that the postulated operations of addition of elements and multiplication of elements by real numbers are continuous in the topology. Cf. [4] and [9]. It suffices to give the system $\mathfrak{U}$ of neighborhoods of the origin. The system of neighborhoods of an arbitrary point a consists of the neighborhoods of the form $U(a) = a + U (U \varepsilon \mathfrak{U})$. A linear topological space E is called *convex*, if the topology of E is defined by means of a system of neighborhoods which are convex open sets, i. e., if $U \varepsilon \mathfrak{U}$, $\alpha, \beta > 0$, $\alpha + \beta = 1$, then $\alpha U + \beta U = U$.[5] We assume also, without loss of generality, the symmetry of $U, -U = U$, for $U \varepsilon \mathfrak{U}$. Hence $b \varepsilon U(a)$ is equivalent with $a \varepsilon U(b)$.

---

[4] $|x| = (\sum_{i=1}^{m} x_i^2)^{1/2}$ for $x \varepsilon E^m$.

In a convex linear topological space E the *pseudonorm* $\| x \|_U (U \varepsilon \mathfrak{U})$ is defined for every $x \varepsilon E$ as follows (cf. [4]):

$$\| x \|_U = \text{g. l. b. } of\ \alpha > 0\ such\ that\ x \varepsilon \alpha U.$$

Then $\| x \|_U$ has the following properties:

(a) $\| x \|_U \geqq 0$,
(b) $\| x + y \|_U \leqq \| x \|_U + \| y \|_U$,
(c) $\| \alpha x \|_U = | \alpha | \, \| x \|_U$ *for any real* α,
(d) $\| x \|_U$ *is a continuous function of* x,
(e) $\| x \|_U < 1$ *is equivalent with* $x \varepsilon U$.

In this paper *we denote by E always a convex linear topological space*, and assume $U(a)$, $\mathfrak{U}$ have the meanings mentioned above.

2. Auxiliary theorems. Before we proceed to the definition of the degree of mapping in E we give some auxiliary theorems. A transformation f of M into E is called *completely continuous* on M if f is continuous on M and the image $f(M)$ of M is a subset of a compact set in E.[6] In this paper we denote by $\mathsf{T}f$ the transformation

$$\mathsf{T}f(x) = x + f(x)$$

for any transformation f in E.

THEOREM 1. *Let M be a closed set in E, and f be completely continuous on M. Then $\mathsf{T}f(M)$ is also closed in E.*

Proof. We have to prove that for any point $a \,\bar{\varepsilon}\, \mathsf{T}f(M)$ there exists a $U \varepsilon \mathfrak{U}$ such that $U(a) \cap \mathsf{T}f(M) = 0$.

For any $p \varepsilon E$ there exists a $U_p \varepsilon \mathfrak{U}$ such that:

$$U_p(p) \cap M = 0 \text{ if } p \,\bar{\varepsilon}\, M, \tag{1}$$

$$U_p(p) \cap (a - f(U_p(p))) = 0 \text{ if } p \varepsilon M, \tag{2}$$

---

[5] By αS (α a real number) we denote the set of all points αx, where $x \varepsilon S$, and by $S_1 + S_2$ the set of all points $x_1 + x_2$, where $x_i \varepsilon S_i$.

[6] This definition is more restricted than the ordinary one, when M is not bounded. We use the term "compact" in the sense of "bicompact."

since $p \neq a - f(p)$ and f is continuous on M. The set $a - f(M)$ is contained in a compact set K, and there exist a finite number of points $p_i \varepsilon K$ $(i = 1, \cdots, k)$ such that

$$\bigcup_{i=1}^{k} 2^{-1} U_{p_i}(p_i) \supset K. \tag{3}$$

Then we shall have $U_0(a) \cap \mathsf{T}f(M) = 0$, if we put

$$U_0 = \bigcap_{i=1}^{k} 2^{-1} U_{p_i}. \tag{4}$$

For if $U_0(a) \cap \mathsf{T}f(M) \neq 0$, then there should exist a $p \varepsilon M$ such that $p + f(p) \varepsilon U_0(a)$, hence $p \varepsilon U_0(a - f(p)) \subset U_0(K)$. Then by (3) and (4) there exists an i such that $a - f(p) \varepsilon U_{p_i}(p_i)$ and $p \varepsilon U_{p_i}(p_i)$. If $p_i \bar{\varepsilon} M$, then by (1) $U_{p_i}(p_i) \cap M = 0$, which contradicts $p \varepsilon M$ and $p \varepsilon U_{p_i}(p_i)$. If $p_i \varepsilon M$, then by (2) $U_{p_i}(p_i) \cap (a - f(U_{p_i}(p_i))) = 0$, which contradicts $a - f(p) \varepsilon U_{p_i}(p_i)$ and $a - f(p) \varepsilon (a - f(U_{p_i}(p_i)))$. Thus the proof is complete.

THEOREM 2. *Let K be a compact set in E. Then for any $U \varepsilon \mathfrak{U}$ there exists a finite-dimensional linear manifold E^m in E and a continuous transformation S of K into E^m such that $S(x) - x \varepsilon U$ for $x \varepsilon K$.*

Proof. By the hypothesis there exist a finite number of points $p_i \varepsilon K$ $(i = 1, \cdots, k)$ such that $K \subset \bigcup_{i=1}^{k} U(p_i)$. Let E^m be the linear manifold spanned by $p_1, \cdots, p_k$. We put $\rho_i(x) = \operatorname{Max}\{(1 - \| x - p_i \|_U), 0\}$ and define S by

$$S(x) = \left(\sum_{i=1}^{k} \rho_i(x)\right)^{-1} \sum_{i=1}^{k} \rho_i(x) p_i.$$

That S satisfies the above statement follows from the convexity of U, the continuity of $\rho_i(x)$, and the equivalence of $\rho_i(x) > 0$ with $x \varepsilon U(p_i)$.

THEOREM 3. *Let G be an open set in an l-dimensional Euclidean space E^l (l finite); f a continuous transformation of $\bar{G}$ into E^l such that $f(x) - x = \phi(x)$ is bounded on G and $\phi(\bar{G}) \subset E^m$, where E^m is an m-dimensional linear manifold in $E^l (m < l)$; and $a \bar{\varepsilon} f(\bar{G} - G)$; then*

$$A^m[a, G^m, f] = A^l[a, G, f],$$

where $G^m = G \cap E^m$ and the superscript of A means that the degree of mapping is considered in the space of the assigned dimension.

Proof. Because the degree of mapping is invariant under a linear transformation of the coordinates, let us take the x_1-axis, $\cdots$, x_m-axis in E^m so that the transformation $x' = f(x)$ shall be of the form

$$x'_i = f_i(x_1, \cdots, x_m, \cdots, x_l) \text{ for } 1 \leqq i \leqq m,$$

$$x'_j = x_j = f_j(x) \text{ for } m + 1 \leqq j \leqq l.$$

Let R be an open sphere with center at the origin and a sufficiently large radius; then, putting $G_0{}^i = R \cap G^i$, we have by v) in § 1,

$$(1) \qquad A^i[a, G^i, f] = A^i[a, G_0{}^i, f] \qquad (i = l, m).$$

Let the distance between a and $f(\bar{G}_0 - G_0)$ be Δ, so that $\Delta > 0$; and let $f_i{}^*(x)$ be functions with continuous partial derivatives on $\bar{G}_0$ such that $\sum_{i=1}^{m} \{f_i{}^*(x) - f_i(x)\}^2 < \Delta^2$ and $f_j{}^*(x) = f_j(x) = x_j$ for $m + 1 \leqq j \leqq l$. Then by the Corollary in § 1,

$$(2) \qquad A^i[a, G_0{}^i, f] = A^i[a, G_0{}^i, f^*] \qquad (i = l, m).$$

Using the notation of [3] we have

$$D(f_1{}^*, \cdots, f_l{}^*/x_1, \cdots, x_l) = D(f_1{}^*, \cdots, f_m{}^*/x_1, \cdots, x_m).^7$$

Thus by the definition of $A[a, G, f^*]$ in [3] we obtain

$$A^l[a, G_0{}^l, f^*] = A^m[a, G_0{}^m, f^*],$$

since all roots of $f^*(x) = a$ are in $E^m (a_j = 0$ for $j > m)$. From this with (1) and (2) the theorem follows.

3. Definition of the degree of mapping in *E*. Let G be an open set in E and f a completely continuous transformation of G into E. Let a be a point not on $\mathrm{T}f(\bar{G} - G)$; then by Theorem 1 there exists a $U \varepsilon \mathfrak{U}$ such that $U(a) \cap \mathrm{T}f(\bar{G} - G) = 0$. By hypothesis there exists a compact set K such that $f(\bar{G}) \subset K$ and by Theorem 2 there exist an m-dimensional linear manifold E^m (m finite) and a continuous transformation S of K into E^m such that $a \varepsilon E^m$, $S(x) - x \varepsilon U$ for $x \varepsilon K$. Then $\mathrm{T}Sf(x) - \mathrm{T}f(x) \varepsilon U$ for $x \varepsilon \bar{G}$; hence $a \bar{\varepsilon} \mathrm{T}Sf(\bar{G} - G)$. Since a linear topological space of finite dimension is linearly homeomorphic to the Euclidean space of the same dimension (cf. [8]), $\mathrm{T}Sf$ transforms $\bar{G}^m = \bar{G} \cap E^m$ into E^m, $\mathrm{T}Sf(x) - x = Sf(x)$ is bounded ($S(K)$ is compact) on $\bar{G}^m$, and $a \bar{\varepsilon} \mathrm{T}Sf(\bar{G}^m - G^m)$ (for $\bar{G}^m - G^m \subset \bar{G} - G$), it follows that $A^m[a, G^m, \mathrm{T}Sf]$, as the degree of mapping

---

[7] $D(f_1, \cdots, f_k/x_1, \cdots, x_k) = \det(\partial f_i/\partial x_j)_{i,j=1,\ldots,k}$.

in E^m, has a definite meaning. Then *we define the degree of mapping* $A[a, G, \mathsf{T}f]$ *in* E *by*

$$A[a, G, \mathsf{T}f] = A^m[a, G^m, \mathsf{T}Sf]. \tag{0}$$

This definition will be legitimized by the following:

THEOREM 4. $A[a, G, \mathsf{T}f]$ *defined by* (0) *is independent of the choice of* E^m *and* S.

Proof. Let $U_i \varepsilon \mathfrak{U}$ $(i = 1, 2)$ be such that

$$U_i(a) \cap \mathsf{T}f(\bar{G} - G) = 0. \tag{1}$$

Let E^m and E^n be linear manifolds of m respectively n dimensions in E, and S_i $(i = 1, 2)$ be continuous transformations of K into E such that

$$S_i(x) - x \,\varepsilon\, U_i \text{ for } x \,\varepsilon\, K,\ a \,\varepsilon\, E^m \cap E^n,\ S_1(K) \subset E^m,\ S_2(K) \subset E^n. \tag{2}$$

By Theorems 1 and 2 we can select a linear manifold E^l of finite dimension l in E and a continuous transformation S_3 of K into E^l such that

$$S_3(x) - x \,\varepsilon\, U_1 \cap U_2 \text{ for } x \,\varepsilon\, K \text{ and } E^m \cup E^n \subset E^l. \tag{3}$$

Then, putting $G^i = G \cap E^i$ $(i = l, m, n)$, we get by Theorem 3

$$A^m[a, G^m, \mathsf{T}S_1f] = A^l[a, G^l, \mathsf{T}S_1f]. \tag{4}$$

If we put $(1 - t)\mathsf{T}S_1f + t\mathsf{T}S_3f = F_t$, then by (2) and (3)

$$F_t(x) - \mathsf{T}f(x) = (1 - t)\{S_1f(x) - f(x)\} + t\{S_3f(x) - f(x)\}.$$
$$\varepsilon\, U_1 \text{ for } x \,\varepsilon\, \bar{G},\quad 0 \leqq t \leqq 1.$$

Thus from (1) and $\bar{G}^l - G^l \subset \bar{G} - G$ it follows that

$$a \,\bar{\varepsilon}\, F_t(\bar{G}^l - G^l) \text{ for } 0 \leqq t \leqq 1. \tag{5}$$

Since $F_t(x) - x = (1 - t)S_1f(x) + S_3f(x)$ is bounded ($S_i(K)$ are compact) for $x \,\varepsilon\, \bar{G}^l$, then by (5) and iv) in § 1, $A^l[a, G^l, F_t]$ is constant for $0 \leqq t \leqq 1$. Thus $A^l[a, G^l, \mathsf{T}S_1f] = A^l[a, G^l, \mathsf{T}S_3f]$. Hence by (4)

$$A^m[a, G^m, \mathsf{T}S_1f] = A^l[a, G^l, \mathsf{T}S_3f].$$

Similarly we get

$$A^n[a, G^n, \mathsf{T}S_2f] = A^l[a, G^l, \mathsf{T}S_3f].$$

nsequently

$$A^m[a, G^m, \mathsf{T}S_1f] = A^n[a, G^n, \mathsf{T}S_2f].$$

letes the proof.

4. Fundamental properties of the degree of mapping in E. The fundamental properties of the degree of mapping stated in § 1 remain valid also for transformations $\mathsf{T}f$ in E. The symbols f, T, K, etc. have the same meaning as in § 3. The validity of i) is evident.

THEOREM 5. *Property* ii) *in* § 1 *is valid for the mapping* $\mathsf{T}f$ (*instead of* f) *in* E, *if* $a \bar{\varepsilon} \mathsf{T}f(\bar{G} - G)$.

Proof. If $a \bar{\varepsilon} \mathsf{T}f(G)$, then $a \bar{\varepsilon} \mathsf{T}f(\bar{G})$. Thus by Theorem 1 there exists a $U \varepsilon \mathfrak{U}$ such that

$$(1) \qquad U(a) \cap \mathsf{T}f(\bar{G}) = 0.$$

Then by Theorem 2 there exists a linear manifold of finite dimension E^m and a continuous transformation S of K into E^m such that

$$(2) \qquad S(x) - x \varepsilon U \text{ for } x \varepsilon K.$$

By the hypothesis and the definition of the degree of mapping in E,

$$0 \neq A[a, G, \mathsf{T}f] = A^m[a, G^m, \mathsf{T}Sf] \quad (G^m = G \cap E^m).$$

Thus by ii) in § 1, for the mapping $\mathsf{T}Sf$ in E^m, there exists a point $x_0 \varepsilon G^m$ such that $a = \mathsf{T}Sf(x_0)$. But by (2), since $f(x_0) \varepsilon K$, $\mathsf{T}Sf(x_0) - \mathsf{T}f(x_0) = Sf(x_0) - f(x_0) \varepsilon U$. Then $\mathsf{T}f(x_0) \varepsilon U(a)$ which contradicts (1).

THEOREM 6. *Property* iii) *in* § 1 *is valid for the mapping* $\mathsf{T}f$ (*instead of* f) *in* E.

Proof. Since $a \varepsilon \mathsf{T}f(\bar{G}_i - G_i)$ $(i = 1, \cdots, k)$, by Theorem 1 there exists a $V \varepsilon \mathfrak{U}$ such that

$$(1) \qquad V(a) \cap \mathsf{T}f(\bar{G}_i - G_i) = 0, \qquad V(a) \cap \mathsf{T}f(\bar{G} - G) = 0,$$

since $\bar{G} - G \subset \bigcup_{i=1}^{k} (\bar{G}_i - G_i)$. By Theorem 2 there exists a finite dimensional linear manifold E^m and a continuous transformation S of K into E^m such that $a \varepsilon E^m$ and

$$(2) \qquad S(x) - x \varepsilon V \text{ for } x \varepsilon K.$$

Then by the definition of the degree of mapping for $\mathsf{T}f$ in E,

$$(3) \qquad \left\{ \begin{array}{l} A[a, G_i, \mathsf{T}f] = A^m[a, G_i^m, \mathsf{T}Sf] \quad (i = 1, \cdots, m), \\ A[a, G, \mathsf{T}f] = A^m[a, G^m, \mathsf{T}Sf]. \end{array} \right.$$

From (1) and (2) also follows

$$(4) \qquad a \,\bar{\varepsilon}\, \mathsf{T}Sf(\bar{G}_i - G_i) \cup \mathsf{T}Sf(\bar{G} - G) \qquad (i = 1, \cdots, m).$$

Let X be the set of all roots of $\mathsf{T}Sf(x) = a$ in $\bar{G}$. Then $X \subset \bar{G}^m$, because from $\mathsf{T}Sf(x) = a$ follows $x = a - Sf(x) \,\varepsilon\, E^m$. From (4) and $\bar{G} = \bigcup_{i=1}^{k} \bar{G}_i$ we obtain $X \subset \bigcup_{i=1}^{k} G_i^m \subset G^m$. Thus by v) for $\mathsf{T}Sf$ (instead of f) in E^m,

$$(5) \qquad A^m[a, G^m, \mathsf{T}Sf] = A^m[\mathrm{a}, \bigcup_{i=1}^{k} G_i^m, \mathsf{T}Sf].$$

Since $G_i^m \cap G_j^m = 0$ $(i \neq j)$, we get by ii) in § 1 for $\mathsf{T}Sf$ (instead of f) in E^m

$$(6) \qquad A^m[a, \bigcup_{i=1}^{k} G_i^m, \mathsf{T}Sf] = \sum_{i=1}^{k} A^m[a, G_i^m, \mathsf{T}Sf].$$

From (3), (5) and (6) we obtain

$$A[a, G, \mathsf{T}f] = \sum_{i=1}^{k} A[a, G_i, \mathsf{T}f].$$

COROLLARY 1. *Property* v) *in* § 1 *is valid for* $\mathsf{T}f$ (*instead of* f) *in* E.

Proof. Put $G - \bar{G}_0 = G$. Then $G \supset G_0 \cup G_1$, $\bar{G} = \bar{G}_0 \cup \bar{G}_1$, $G_0 \cap G_1 = 0$. Apply Theorem 6 and the relation $A[a, G_1, \mathsf{T}f] = 0$ (by Thoerm 5).

Property iv) in § 1 becomes as follows:

THEOREM 7. *Let* f_t *be a transformation of* $\bar{G}$ *into* E *such that* $f_t(x)$ *is a continuous function of* (t, x) *for* $0 \leqq t \leqq 1$, $x \,\varepsilon\, \bar{G}$ *and is always contained in a compact set* K.[8] *If* $a(t)\,(\varepsilon\, E)$ *is continuous for* $0 \leqq t \leqq 1$ *and* $a(t) \,\bar{\varepsilon}\, \mathsf{T}f_t(\bar{G} - G)$ *for* $0 \leqq t \leqq 1$, *then* $A[a(t), G, \mathsf{T}f_t]$ *is constant for* $0 \leqq t \leqq 1$.

To prove this we use the following

LEMMA 1. *Let* K_i $(i = 1, 2)$ *be compact sets in* E; *then* $K_1 + K_2$[5] *is also compact. Let* K *be a compact set in* E; *then the set* $\hat{K}$ *of all points* tx *such that* $0 \leqq t \leqq 1$, $x \,\varepsilon\, K$, *is also compact.*

Proof. $K_1 \times K_2$ is a compact set in $E \times E$; then by the continuous mapping $x' = x_1 + x_2 (x_i \,\varepsilon\, K_i)$, $K_1 \times K_2$ goes onto the compact set $K_1 + K_2$

---

[8] This condition is weaker than that $f_t(x)$ is completely continuous for $x \,\varepsilon\, G$ and uniformly continuous in t for $x \,\varepsilon\, \bar{G}$, $0 \leqq t \leqq 1$.

in E. Similarly $K \times \langle 0, 1\rangle$ is a compact set in $E \times E^1$, which is mapped by $x' = tx$ onto the compact set $\hat{K}$ in E.

Proof of Theorem 7. First we assume that $a(t)$ is constant and put $a(t) = a$. Let τ be any fixed value of t from $0 \leqq t \leqq 1$. Consider the product space $E^* = E \times E^1$ (E^1 is the 1-dimensional space $-\infty < t < +\infty$), which is also a convex linear topological space with a system of neighborhoods of the origin, $\mathfrak{U}^* = \{U \times (-\delta, \delta) \mid U \varepsilon \mathfrak{U}, \delta > 0\}$. Let G^* be the open set in E^* defined by $G^* = G \times (-\infty, +\infty)$. Let us define the transformation f^* in E^* by $f^*(x, t) = (f_{\langle t\rangle}(x), 0)$, where $\langle t\rangle = 0$ for $t < 0$, $\langle t\rangle = t$ for $0 \leqq t \leqq 1$, $\langle t\rangle = 1$ for $t > 1$. Then $f^*(\bar{G}^*) \subset (K, 0) = K^*$ (a compact set in E^*). Since $\bar{G}^* - G^* = (\bar{G} - G) \times (-\infty, +\infty)$, then $(a, \tau) \bar{\varepsilon} \mathsf{T}f^*(\bar{G}^* - G^*)$. Thus by Theorem 1 there exists a $U^* \subset \mathfrak{U}^*$ such that

$$U^*(a, \tau) \cap \mathsf{T}f^*(\bar{G}^* - G^*) = 0.$$

This means that there exists a $\delta > 0$ and a $U \varepsilon \mathfrak{U}$ such that

$$U(a) \cap \mathsf{T}f_t(\bar{G} - G) = 0 \text{ for } |t - \tau| < \delta. \tag{1}$$

By Theorem 2 there exists a finite-dimensional linear manifold E^m and a continuous transformation S of K into E^m such that $a \varepsilon E^m$ and $S(x) - x \varepsilon U$ for $x \varepsilon K$. Then by (1) and the definition of the degree of mapping for $\mathsf{T}f_t$ in E,

$$A[a, G, \mathsf{T}f_t] = A^m[a, G^m, \mathsf{T}Sf_t] \text{ for } |t - \tau| < \delta. \tag{2}$$

From (1) and $\mathsf{T}f_t(x) - \mathsf{T}Sf_t(x) \varepsilon U$ follows $a \bar{\varepsilon} \mathsf{T}Sf_t(\bar{G}^m - G^m)$ for $|t - \tau| < \delta$. Thus by iv) in § 1, $A^m[a, G^m, \mathsf{T}Sf_t]$, and hence by (2) $A[a, G, \mathsf{T}f_t]$, is constant for $|t - \tau| < \delta$. Then by using the covering theorem for the closed interval $0 \leqq t \leqq 1$ we obtain the theorem for constant $a(t)$.

The constancy of $a(t)$ can be dropped, if we consider $f_t(x) - a(t)$ instead of $f_t(x)$, since the set $K + \{-a(t) \mid 0 \leqq t \leqq 1\}$ is also compact by Lemma 1. For, $A[a, G, \mathsf{T}f] = A[0, G, \mathsf{T}f - a]$ if $a \bar{\varepsilon} \mathsf{T}f(\bar{G} - G)$.[9]

COROLLARY 2. *Let f be a completely continuous transformation of $\bar{G}$ into E, where G is an open set in E. If a is a point of E and there exists*

---

[9] This relation is valid if E is finite-dimensional (put $F_t = Tf - ta$, $a(t) = (1 - t)a$ and apply iv) in §1). For the general case we have by definition $A[a, G, Tf] = A^m[a, G^m, TSf]$, $A[0, G, Tf - a] = A^m[0, G^m, TSf - a]$, where A^m, G^m and S have the same meaning as in §2. Hence the relation holds for general E.

a $U \varepsilon \mathfrak{U}$ *such that* $U(a) \cap \mathsf{T}f(\bar{G}-G)=0$, *then for every point* $a' \varepsilon U(a)$, $A[a', G, \mathsf{T}f]=A[a, G, \mathsf{T}f]$.

Proof. Put $a(t)=(1-t)a+ta'$ and apply Theorem 7.

COROLLARY 3. *Let* f, G, a *and* U *satisfy the same condition as in Corolary 2. Then for every completely continuous transformation* f' *of* $\bar{G}$ *into* E *such that*

$$f'(x)-f(x) \varepsilon U \text{ for } x \varepsilon \bar{G},$$

$$A[a, G, \mathsf{T}f']=A[a, G, \mathsf{T}f].$$

Proof. Put $f_t=(1-t)f+tf'$ and apply Theorem 7. $f_t(\bar{G})$ is contained in a definite compact set for $0 \leqq t \leqq 1$ by Lemma 1.

5. Product theorem. For the composition of two transformations in a finite dimensional E^m the following holds:

THEOREM 8. *Let* G *be an open set in* E^m *and* f *a continuous mapping of* $\bar{G}$ *into* E^m *such that* $f(x)-x$ *is bounded on* G. *Let* H *be an open set containing* $f(\bar{G})$ *and* H_i $(i=1,2,\cdots)$ *the components of the open set* $H-f(\bar{G}-G)$. *Let* ϕ *be a continuous mapping of* $\bar{H}$ *into* E^m *such that* $\phi(x)-x$ *is bounded, and* a *a point of* E^m *such that* $a \bar{\varepsilon} \phi f(\bar{G}-G) \cup \phi(\bar{H}-H)$. *Then we have*

$$A[a, G, \phi f]=\sum_i A[a, H_i, \phi] \cdot A[b_i, G, f],$$

where b_i *is an arbitrary point of* H_i.

Proof. Cf. [3].

This theorem can also be extended to the general case of E as follows:

THEOREM 9. *Let* G *be an open set in* E *and* f *a completely continuous mapping of* $\bar{G}$ *into* E. *Let* H *be an open set containing* $\mathsf{T}f(\bar{G})$ *and* H_i *the components of the open set* $H-\mathsf{T}f(\bar{G}-G)$. *Let* ϕ *be completely continuous mapping of* $\bar{H}$ *into* E *and* a *a point of* E *such that*

$$a \bar{\varepsilon} \mathsf{T}\phi{*}f(\bar{G}-G) \cup \mathsf{T}\phi(\bar{H}-H),$$

where

$$\mathsf{T}\phi{*}f=\mathsf{T}\phi\mathsf{T}f \cdot \ (\phi{*}f=f+\phi+\phi\mathsf{T}f).$$

Then we have

$$A[a, G, \mathsf{T}\phi{*}f]=\sum_i A[a, H_i, \mathsf{T}\phi] \cdot A[b_i, G, \mathsf{T}f],$$

where b_i *is an arbitrary point of* H_i.

Remarks. (i) By the hypothesis there exist compact sets K_1 and K_2 such that $f(\bar{G}) \subset K_1$ and $\phi(\bar{H}) \subset K_2$. Since $\phi{*}f(\bar{G}) \subset f(\bar{G}) + \phi(\bar{H}) \subset K_1 + K_2$, then $\phi{*}f$ is also completely continuous on G.

(ii) The set of all roots of $\mathsf{T}\phi(x) = a$ is compact (a closed subset of $a - K_2$). Hence there are only a finite number of H_i such that $A[a, H_i, \mathsf{T}\phi] \neq 0$; thus the summation $\sum_i$ is to be taken only over those i such that $A[a, H_i, \mathsf{T}\phi] \neq 0$.

(iii) Since any two points b' and b'' in the same H_i can be joined by a polygonal line without touching $\mathsf{T}f(\bar{G} - G)$, it follows that $A[x, G, \mathsf{T}f]$ is constant in each H_i.

Proof of Theorem 9. Let D_k be the set of points $p \varepsilon H$ such that $A[p, G, \mathsf{T}f] = k$. Then we have

$$\bar{D}_k - D_k \subset (\bar{H} - H) \cup \mathsf{T}f(\bar{G} - G), \tag{1}$$

and

$$\sum_i A[a, H_i, \mathsf{T}\phi] \cdot A[b_i, G, \mathsf{T}f] = \sum_k A[a, D_k, \mathsf{T}\phi] \cdot k.$$

Thus we have to prove

$$A[a, G, \mathsf{T}\phi{*}f] = \sum_k A[a, D_k, \mathsf{T}\phi] \cdot k. \tag{*}$$

By the hypothesis and Theorem 1 there exists a $V \varepsilon \mathfrak{U}$ such that

$$V(a) \cap (\mathsf{T}\phi(\bar{H} - H) \cup \mathsf{T}\phi{*}f(\bar{G} - G)) = 0. \tag{2}$$

From (1) and (2),

$$V(a) \cap \mathsf{T}\phi(\bar{D}_k - D_k) = 0. \tag{3}$$

By Theorem 2 there exists a finite-dimensional linear manifold E^n and a continuous transformation S_2 of K_2 (Remark (i)) into E^n such that $S_2(x) - x \varepsilon V$ for $x \varepsilon K_2$. Thus

$$\mathsf{T}S_2\phi(x) - \mathsf{T}\phi(x) \varepsilon V \text{ for } x \varepsilon \bar{H}. \tag{4}$$

Then by Corollary 3 we obtain from (3) and (4)

$$A[a, D_k, \mathsf{T}\phi] = A[a, D_k, \mathsf{T}S_2\phi]. \tag{5}$$

Now let X be the set of all roots of $\mathsf{T}S_2\phi(x) = a$. Then X is compact since $X \subset a - S_2(K_2)$. From (2) and (4) it follows that $a \bar{\varepsilon} \mathsf{T}S_2\phi\mathsf{T}f(\bar{G} - G)$.

Thus $X \cap \mathsf{T}f(\bar{G}-G)=0$. Then because X is compact and $\mathsf{T}f(\bar{G}-G)$ is closed, there exists a $U \varepsilon \mathfrak{U}$ such that

$$U(X)\cap \mathsf{T}f(\bar{G}-G)=0. \tag{6}$$

By Theorem 2 there exists finite-dimensional linear manifold E^m and a continuous transformation S_1 of K_1 into E^m such that $a \varepsilon E^m$ and $S_1(x)-x \varepsilon U$ for $x \varepsilon K_1$. Hence

$$\mathsf{T}S_1f(x)-\mathsf{T}f(x)\ \varepsilon\ U \quad \text{for} \quad x \varepsilon \bar{G}. \tag{7}$$

Then by (6) and Corollary 3,

$$A[p, G, \mathsf{T}f]=A[p, G, \mathsf{T}S_1f] \quad \text{for} \quad p \varepsilon X. \tag{8}$$

Now let D'_k be the set of points $p \varepsilon H$ such that $A[p, G, \mathsf{T}S_1f]=k$. Then by (8) the set $X \cap D_k$ of all roots of $\mathsf{T}S_2\phi(x)=a$ in D_k coincides with $X \cap D'_k$, that of all roots in D'_k. Thus by Corollary 1,

$$A[a, D_k, \mathsf{T}S_2\phi]=A[a, D'_k, \mathsf{T}S_2\phi] \quad (=A[a, D_k \cap D'_k, \mathsf{T}S_2\phi]).$$

Then by (5)

$$A[a, D_k, \mathsf{T}\phi]=A[a, D'_k, \mathsf{T}S_2\phi]. \tag{9}$$

On the other hand, from (2), (4), and $\mathsf{T}f(\bar{G})\subset H$, we get by Corollary 3,

$$A[a, G, \mathsf{T}\phi^*f]=A[a, G, \mathsf{T}(S_2\phi)^*f]. \tag{10}$$

Now putting $(1-t)f+tS_1f=f_t$, we have $\mathsf{T}S_2\phi\mathsf{T}f_t=\mathsf{T}(S_2\phi)^*f_t$, where $(S_2\phi)^*f_t=(1-t)f+tS_1f+S_2\phi(\mathsf{T}f_t)$. Thus

$$(S_2\phi)^*f_t(\bar{G})\subset \hat{K}_1+\widehat{S_1(\hat{K}_1)}+S_2(K_2)=K_3, \tag{11}$$

where K_3 is a compact set by Lemma 1. From (7) we get $\mathsf{T}f_t(x)-\mathsf{T}f(x)$ $=t(S_1f(x)-f(x))\ \varepsilon\ U$ for $0\leqq t\leqq 1$. Thus by (6), $p\ \bar{\varepsilon}\ \mathsf{T}f_t(\bar{G}-G)$ for $p \varepsilon X$, $0\leqq t\leqq 1$. Hence $a\ \bar{\varepsilon}\ \mathsf{T}(S_2\phi)^*f_t(\bar{G}-G)$ for $0\leqq t\leqq 1$. Thus by (11) and Theorem 7, $A[a, G, \mathsf{T}(S_2\phi)^*f_t]$ is constant for $0\leqq t\leqq 1$; hence $A[a, G, \mathsf{T}(S_2\phi)^*f]=A[a, G, \mathsf{T}(S_2\phi)^*(S_1f)]$, and by (10),

$$A[a, G, \mathsf{T}\phi^*f]=A[a, G, \mathsf{T}(S_2\phi)^*(S_1f)]. \tag{12}$$

Now let E^l be a finite-dimensional linear manifold containing both E^m and E^n. Then $S_2\phi(\bar{G})\subset E^l$, $(S_2\phi)^*(S_1f)(\bar{G})\subset E^l$, and by the definition of the degree of mapping in E

$$\begin{cases} A[a, D'_k, \mathsf{T}S_2\phi]=A^l[a, D'^l_k, \mathsf{T}S_2\phi], \\ A[a, G, \mathsf{T}(S_2\phi)^*(S_1f)]=A^l[a, G^l, \mathsf{T}(S_2\phi)^*(S_1f)]. \end{cases} \tag{13}$$

But, since $D'^{\,l}_k$ is just the set of points $p \,\varepsilon\, H$ such that $A^l[p, G^l, \mathrm{T}S_1 f] = k$, we obtain by Theorem 8

$$A^l[a, G^l, \mathrm{T}(S_2\phi)^*(S_1 f)] = \sum_k A^l[a, D'^{\,l}_k, \mathrm{T}S_2\phi] \cdot k, \tag{14}$$

using considerations similar to those in the first part of this proof Then from (9), (12), (13) and (14) follows (*).

6. Invariance of domain in complete metric *E*.

THEOREM 10. *Let E be complete metric, G an open set in E, and f a completely continuous transformation of $\bar{G}$ into E. If $\mathrm{T}f$ affords a one-one correspondence between $\bar{G}$ and $\mathrm{T}f(\bar{G})$, then $\mathrm{T}f(G)$ is an open set in E, and $A[b, G, \mathrm{T}f] = \pm 1$ for any $b \,\varepsilon\, \mathrm{T}f(G)$.*

To prove this theorem we shall first give two lemmas.

LEMMA 2. *If K is a compact set in a complete metric E, then the smallest closed convex set $\hat{K}$ containing K is also compact.*

Proof. Using the convexity of $\mathfrak{U}$ one can easily prove that $\hat{K}$ is totally bounded[10]; hence the closed set $\hat{K}$ is compact (in complete metric E).

LEMMA 3.[11] *If $f(x)$ is continuous on a closed set M in a separable complete metric E, and $f(M)$ is contained in a compact convex set K in E, then $f(x)$ can be extended to a continuous function on the whole of E in such a way that $f(E) \subset K$.*

Proof. Let $\{a_n\}$ be a denumerable set dense on M. We put

$$\rho(p, x) = \mathrm{Max}\{(2 - \mathrm{dist}(p, x)/\mathrm{dist}(p, M)), 0\} \text{ for } p \,\bar{\varepsilon}\, M;$$

then $\sum_{n=1}^{\infty} 2^{-n}\rho(p, a_n) f(a_n)$ converges uniformly on $E - M$,[12] and we define $f^*(p)$ by

$$f^*(p) = \Big(\sum_{n=1}^{\infty} 2^{-n}\rho(p, a_n)\Big)^{-1} \sum_{n=1}^{\infty} 2^{-n}\rho(p, a_n) f(a_n) \text{ for } p \,\bar{\varepsilon}\, M,$$
$$= f(p) \text{ for } p \,\varepsilon\, M.$$

Then $f^*(p)$ is an extension of $f(p)$ as desired.

---

[10] For any $U \,\varepsilon\, \mathfrak{U}$ there exists a finite number of $p_i \,\varepsilon\, \hat{K}$ $(i = 1, \cdots, k)$ such that $\hat{K} \subset \bigcup_{i=1}^{k} U(p_i)$.

[11] This lemma I owe to Professor S. Kakutani.

[12] Because $\sum_{n=i}^{j} 2^{-n}\rho(p, a_n) f(a_n) \,\varepsilon\, 2^{-i+1}K$ (we assume $0 \,\varepsilon\, K$); hence for sufficiently large N, $\sum_{n=i}^{j} 2^{-n}\rho(p, a_n) f(a_n) \,\varepsilon\, U$ ($U \,\varepsilon\, \mathfrak{U}$ is arbitrarily given) if $i > N$, since K is compact.

Proof of Theorem 10. Let $b = \mathsf{T}f(a)$ be any point of $\mathsf{T}f(G)$, and let $f(\bar{G})$ be contained in a compact set K. By Lemma 2 we can assume that K is convex. Let E_0 be the smallest closed linear manifold containing K and a. Then E_0 is a separable complete metric subspace of E. Put $G_0 = G \cap E_0$, then $\bar{G}_0$ is transformed by $\mathsf{T}f$ into E_0. Then we have, since $b \,\bar{\varepsilon}\, \mathsf{T}f(\bar{G} - G)$, by the manner of definition of the degree of mapping,

$$(1) \qquad A[b, G, \mathsf{T}f] = A[b, G_0, \mathsf{T}f] \text{ in } E_0.$$

Now let us confine ourselves to E_0. Put $\mathsf{T}f(\bar{G}_0) = M$; then M is closed by Theorem 1. The inverse mapping $(\mathsf{T}f)^{-1}$ of M onto $\bar{G}_0$ is also continuous, because every closed subset of $\bar{G}_0$ corresponds to a closed subset of M by $(\mathsf{T}f)^{-1}$ (Theorem 1). Now put $x' = \mathsf{T}f(x)$, $(\mathsf{T}f)^{-1}(x') - x' = \phi_0(x')$; then $\phi_0(x') = -f(x)$, and hence $\phi_0(M) = -f(\bar{G}_0) \subset -K$. Therefore $\phi_0(x)$ is completely continuous on M, and $\mathsf{T}\phi_0 = (\mathsf{T}f)^{-1}$. Since M is closed in a separable complete metric E_0, and $\phi_0(M) \subset -K$ (a compact convex set in E_0), then by Lemma 3, $\phi_0(x)$ can be extended to a continuous $\phi(x)$ on the whole of E_0 in such a way that $\phi(E_0) \subset -K$.

Let H_i be the components of the open set $E_0 - \mathsf{T}f(\bar{G}_0 - G_0)$. Since $\mathsf{T}\phi{*}f(x) = \mathsf{T}\phi_0\mathsf{T}f(x) = x$ for $x \,\varepsilon\, \bar{G}_0$, then $a \,\bar{\varepsilon}\, \mathsf{T}\phi{*}f(\bar{G}_0 - G_0)$. Thus by Theorem 9 we obtain, since $\bar{E}_0 - E_0 = 0$ and $A[a, G_0, \mathsf{T}\phi{*}f] = 1$,

$$(2) \qquad 1 = \sum_i A[a, H_i, \mathsf{T}\phi] \cdot A[b_i, G_0, \mathsf{T}f],$$

where b_i is an arbitrary point of H_i. Consequently there exists an $i = k$ such that

$$(3) \qquad A[a, H_k, \mathsf{T}\phi] \cdot A[b_k, G_0, \mathsf{T}f] \neq 0.$$

Then by Theorem 5 there exists a $b_k \,\varepsilon\, H_k$ such that $a = \mathsf{T}\phi(b_k)$ and $b_k \,\varepsilon\, \mathsf{T}f(G_0) \subset M$. Thus $a = \mathsf{T}\phi_0(b_k) = (\mathsf{T}f)^{-1}(b_k)$; hence $b_k = \mathsf{T}f(a) = b$. Since there is only one H_k such that $b \,\varepsilon\, H_k$, we can assume $k = 1$ and $A[a, H_i, \mathsf{T}\phi] \cdot A[b_i, G_0, \mathsf{T}f] = 0$ for all $i > 1$. Thus from (2) follows

$$A[a, H_1, \mathsf{T}\phi] \cdot A[b, G_0, \mathsf{T}f] = 1.$$

Since the factors of the left side of this equation must be integers we get $A[b, G_0, \mathsf{T}f] = \pm 1$. Therefore from (1), returning to E,

$$A[b, G, \mathsf{T}f] = \pm 1.$$

Since $b \,\bar{\varepsilon}\, \mathsf{T}f(\bar{G} - G)$, there exists a $U \,\varepsilon\, \mathfrak{U}$ such that $U(b) \cap \mathsf{T}f(\bar{G} - G) = 0$.

Then by Corollary 2, $A[b', G, \mathsf{T}f] = \pm 1$ for any $b' \varepsilon U(b)$. Hence by Theorem 5, $U(b) \subset \mathsf{T}f(G)$. Thus the proof is complete.

OSAKA UNIVERSITY,
OSAKA, JAPAN.

REFERENCES.

[1] J. Leray and J. Schauder, "Topologie et équations fonctionelles," *Annales Scientifiques de l'École Normale Supérieure*, vol. 51 (1934), pp. 45-78.

[2] ———, "Topologie des espaces abstraits de M. Banach," *Comptes Rendus de l'Académie des Sciences* (Paris), vol. 200 (1935), pp. 1082-1084.

[3] M. Nagumo, "A theory of the degree of mapping based on infinitesimal analysis," *American Journal of Mathematics*, vol. 73 (1951), pp. 485–496.

[4] J. von Neumann, "On complete topological spaces," *Transactions of the American Mathematical Society*, vol. 37 (1935), pp. 1-20.

[5] E. Rothe, "The theory of topological order in some linear topological spaces," *Iowa State College Journal of Science*, vol. 13 (1939), pp. 373-390.

[6] J. Schauder, "Invarianz des Gebietes in Funktionalräumen," *Studia Mathematica*, vol. 1 (1929), 123-139.

[7] ———, "Über den Zusammenhang zwischen der Eindeutigkeit und Lösbarkeit partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus," *Mathematische Annalen*, vol. 106 (1932), pp. 661-721.

[8] A. Tychonoff, 'Ein Fixpunktsatz," *Mathematische Annalen*, vol. 111 (1935), pp. 767-776.

[9] J. Wehausen, "Transformations in linear topological spaces," *Duke Mathematical Journal*, vol. 4 (1938), pp. 157-169.

THE TOPOLOGY OF THE LEVEL CURVES OF HARMONIC FUNCTIONS WITH CRITICAL POINTS.* [1]

By William M. Boothby.

Introduction. In a previous paper,[2] of which this is a continuation, topological properties of curve families which filled the Euclidean plane π, or a simply connected domain in π, were investigated. The families were assumed regular (i. e. locally homeomorphic to parallel lines) except at a possibly infinite collection of isolated singularities at each of which the family had the structure of a multiple saddle point; such families were called *branched regular curve families.* Further investigation of these families, in particular their relation to harmonic functions, is the aim of this paper. In what follows the definitions and theorems in [I] will be assumed, and the same notation will be used. In particular F, G will denote branched regular curve families filling the plane π, B will denote the set of singular points, R the domain $\pi - B$ in which F is regular, and so on. The Euclidean plane will be taken as a model for all simply connected domains.

The principal result of [I] was to prove that any branched regular curve family F filling π can be given as the family of level curves of a function $f(p)$ which is continuous on all of π and has no relative extrema. This generalizes a portion of [II] in which the same theorem is proved for a curve family without singularities in π. In this paper there are two main results: the first, proved in Section **1**, is that F is actually homeomorphic to the level curves of a harmonic function; the second, proved in Section **2**, asserts the existence of a decomposition of F into a countable collection of subfamilies of curves, each of which has the structure of the parallel lines $y = \text{constant}$ of the upper half-plane. Such subfamilies will be called half-parallel, and this decomposition has consequences for the study of harmonic functions and analytic functions which will be mentioned below. These two results generalize

---

* Received January 30, 1950.

[1] The material in this paper and the preceding paper was taken from the author's Ph. D. thesis at the University of Michigan. The author wishes to express his gratitude to Professor Wilfred Kaplan for his guidance in this research and his advice in the preparation of this paper.

[2] The Topology of Regular Curve Families with Multiple Saddle Points (pp. 405-438 of this volume). Theorems or section numbers preceded by I refer to this paper.

those of Kaplan [III] to curve families with singularities of the saddle point type.

The methods of proof used in [I] will also be of value here, i. e., it was first noted that the family F decomposes into single curves extending to infinity in each direction and collections of curves ending at branch points, where each collection forms a tree. Then it was shown that by removing enough curves we are left with a simply connected domain R^* in which the remaining curves F^* of F form a regular family. This allows us to use the theorems of Kaplan. By this method we are able to show in **1. 2** that for any branched regular curve family F, there exists a complementary family G, i. e. there is a branched regular curve family G with the same singularities as F and such that no pair of curves of F and G intersect more than once. For, from Kaplan [IV] it follows that for the family F^* filling R^* there is such a family G^*. Thus it remains only to modify G^* (along the curves removed from π to give us R^*) in such a fashion that the modified family $\tilde{G}^*$ becomes complementary to F when we replace the curves again. The details of this procedure are elaborated in **1. 2**.

Now given F, G complementary, by [I] there exist two functions f, g defined on π which have F, G as level curves. Using f and g, we may define a map $T:\pi \rightarrow uv$-plane by $T(p) = [f(p), g(p)]$ and since f, g have no extrema and have regular curve families as level curves it may easily be shown that T is light and interior. It follows from well known theorems that there is a homeomorphism $h: D \rightarrow \pi$, D a simply connected domain in the xy-plane, such that $w(x,y) = T[h(x,y)]$ is harmonic; but h maps the level curves of w onto F. This concludes in outline the proof of the first principal theorem. Since the converse is well known, this theorem give a characterization by *local* topological properties of the level curves of a function harmonic in a simply connected domain. We note an important but immediate corollary: F is homeomorphic to the family of solutions of a system of differential equations: $dy/dt = p(x,y)$, $dx/dt = q(x,y)$. Section **1. 3** gives us in detail the proof outlined above.

In Section **2** the decomposition theorem mentioned is proved, again by a reduction to theorems of Kaplan for families without singularities in a simply connected domain, in this case, as above, applied to the family F^* filling, and regular everywhere, in the domain R^* obtained by removing curves from F. The importance of this theorem lies in its possible applications to functions analytic in a simply connected domain, as follows. For example, let $u(x,y)$ be the real part of an entire function. The family F of its level curves will then be a branched regular curve family. Now for F by the decomposition theorem there exists a countable set A and a decomposition of F into sets D_α, $\alpha \in A$, which are simply connected, overlap at most on their boundaries (and then along curves of F) and in each of which our entire function is 1-1. Thus the sets D_α furnish a decomposition of the Riemann surface of the inverse function into sheets. This is again a generalization of Kaplan [III] where these results were obtained for analytic functions, with non-vanishing derivative, defined in a simply connected domain.

1. The Branched Regular Curve Family as the Level Curves of a Harmonic Function.

1.1. Preliminary properties and definitions. In this paper a slightly more general definition of cross-section will be needed than in [I], as follows: a curve in R is a cross-section if every arc on it is a cross-section. This removes the restriction that a cross-section be an arc, i. e. it may extend to infinity in one or both directions or even be a bounded open or half-open curve.

Whitney [VI] has shown that in an orientable regular curve family filling a region S there is a function $f(p,t)$ with the properties: for each p in S and any t in $-\infty < t < \infty$, there is a unique point $q = f(p,t)$ lying on the curve C through p; $f(p,t)$ is continuous in both variables; $f(p,0) = p$ and as t increases (decreases) $f(p,t)$ moves continuously in the positive (negative) direction on C. Just as in [II] we have as an immediate corollary to this theorem the following:

THEOREM 1.1-1. *Let γ be a cross-section in F and $S(\gamma)$ the set of curves of F crossing γ. Then $S(\gamma)$ forms an open, simply connected set, F is regular in this set and, in fact there is a homeomorphism of $S(\gamma)$ onto* (i) *a strip $0 \leqq y \leqq 1$,* (ii) *a half-plane $y \geqq 0$ or* (iii) *the xy-plane, carrying the curves of F in $S(\gamma)$ onto the lines $y =$ constant, the case depending on whether* (i) *γ is an arc,* (ii) *γ is a half-open curve, or* (iii) *γ is an open curve, respectively.*

Proof. We will prove only case (ii), the proofs of the other cases being similar. It is a consequence of Theorem I 1.2-2 that $S(\gamma)$ is open and F is regular in $S(\gamma)$. From what follows it is clear that the domain filled by $S(\gamma)$ is simply connected, since it is homeomorphic to a half-plane. Let γ be parametrized by τ, i. e. $\phi(\tau)$ maps $0 \leqq \tau < \infty$ homeomorphically onto γ. Now $S(\gamma) - \gamma$ consists of two disjoint domains A and B and each curve, since it crosses γ exactly once, has an arc in A and an arc in B. Hence the curve family $F[S(\gamma)]$ is orientable, since we may assign to each C in this set the direction as positive along which we pass from A into B. Then by the above mentioned theorem of Whitney we have a function $f(p, t)$ defined on S. Let $p = \phi(\tau)$, then the function $f(\phi(\tau), t)$ at once gives us, if we set $y = \tau$, $x = t$, a homeomorphism of $S(\gamma)$ onto the upper half plane with $x = 0$ the image of γ. This is essentially the same as the situation in [II, p. 174, Theorem 30].

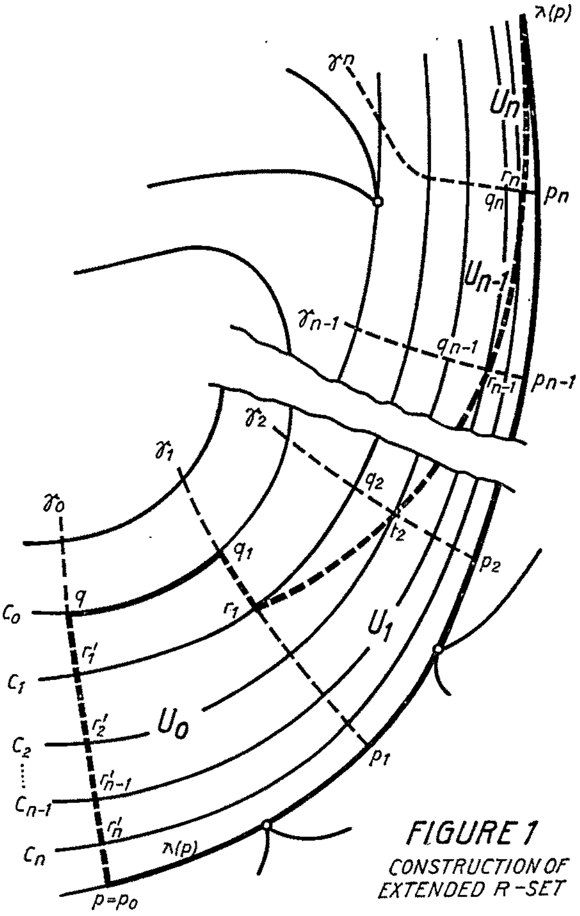


FIGURE 1
CONSTRUCTION OF EXTENDED R-SET

In [I] a useful "half-neighborhood" called an r-set was defined for any arc pq lying on a "maximal curve" C^*. The set was, briefly, a closed set in $\mathcal{D}^*(C^*)$ which abutted on pq and in which the curve family had the structure of the family F_1 of parallel lines, $y = k$, in a closed rectangle $R_1 = \{(x, y) \mid |x| \leqq 1, 0 \leqq y \leqq 1\}$, the only singular points in this set being those on pq itself. Below we shall need a similar half-neighborhood for a half-open curve $\lambda(p)$ extending from a point p to infinity along a maximal curve C^*. Of course, the most important case of such a situation is the "cut," $\lambda(b)$, from a branch point b. An *extended r-set*, $U(\lambda(p))$, of $\lambda(p)$ on C^* will then be a closed set contained in $\mathfrak{D}^*(C^*)$ together with a definite homeomorphism k of this closed set onto $\tilde{R}_1 = R_1 - \{(1,0)\}$, i. e. R_1 without its lower, left-hand corner point, where the following conditions are satisfied by U and k: (1) $F[U]$ is homeomorphic to F_1; (2) the inverse image under k of $x = -1$ is a cross-section, which we will denote by γ, joining p to a point q, $k(p) = (-1, 0)$ and $k(q) = (-1, -1)$; (3) the inverse image of $y = 1$ is an arc from q to $q_1 = k^{-1}[(1,1)]$ on a curve of F; and (4) the inverse image of $x = +1$ is a half-open cross-section Γ from q_1 to infinity, asymptotic to $\lambda(p)$; finally (5) k maps $\lambda(p)$ on the portion of $y = 0$ in $\tilde{R}_1$. Of course, the missing corner-point $(1, 0)$ of $\tilde{R}_1$ corresponds to the point at infinity. (See Figure 1.)

THEOREM 1. 1-2. *Let $\lambda(p)$ be a half-open arc on C^* as defined above and let $V[\lambda(p)]$ be any open set containing $\lambda(p)$, then there is an extended r-set, $U(\lambda(p))$ interior to V.*

Proof. The proof will consist of two parts: the first part (A) being the description of the set U which is to be the extended r-set, and the second part (B) being the description of the homeomorphism k from U to $\tilde{R}_1$. (Fig. 1.)

(A) We begin by choosing a sequence $p = p_0, p_1, p_2, \cdots$ of regular points on $\lambda(p)$ which recede monotonely to infinity along that half-open curve. We shall denote the arc joining p_n to p_{n+1} on $\lambda(p)$ merely by p_np_{n+1}, and as first step we shall choose rather carefully an r-set U_n abutting on p_np_{n+1} for each n. From an examination of the discussion of r-sets in [I], it is not difficult to see that given an arc on a maximal chain whose two endpoints are regular points, if we take arbitrary cross-sections through these endpoints, then it is always possible to find an r-set abutting on this arc whose cross-sectional sides are on these chosen cross-sections. Now let us choose for each n a cross-section γ_n extending from p_n into $\mathfrak{D}^*(C^*)$. For each n we shall choose the r-set U_n interior to V and complying with the following conditions: first, so that its cross-sectional sides are on γ_n and γ_{n+1}, and so that it lies within an ϵ_n-neighborhood of p_np_{n+1}, $\epsilon_n \to 0$. And second, having chosen U_{n-1}, and denoting by q_np_n the arc on γ_n which forms that one of the cross-sectional sides of U_{n-1} on γ_n, we choose U_n in such a manner that its cross-sectional side r_np_n along γ_n is contained in q_np_n, i. e. r_n lies between q_n and p_n on γ_n. Then $U_{n-1} \cap U_n = r_np_n$ by Theorem I 3. 2-2, and C_{r_n} (denoted below by C_n) intersects γ_0 at some point r'_n. Of course, each U_n is in $\mathfrak{D}^*(C^*)$ since each γ_n is. Thirdly, we require the U_n so chosen that r_n is contained in a

δ_n-neighborhood of $p = p_0$ for a definite sequence $\delta_n \to 0$. The properties of r-sets as developed in [I] make it obvious that these conditions on U_n can be complied with. (Note: below we denote γ_0 simply by γ.)

Let k_n denote the homeomorphism of U_n onto R_1; we shall always assume k_n so chosen that as we move from p towards infinity on $\lambda(p)$ the image point moves from left to right along the x-axis. Then if we consider the image of U_n in R_1, we have $k_n(p_n) = (-1, 0)$, $k_n(r_n) = (-1, -1)$, $k_n(p_n r_n) = (\text{line } x = -1)$; $k_n(p_{n+1}) = (1, 0)$, $k_n(q_{n+1}) = (1, 1)$ and $k_n(r_{n+1}) = (1, a)$ where $0 < a < 1$. Now it is clear, as noted above, from the description of the sets U_n, that the curves C_n determined by r_n and q_{n+1}, $n = 1, 2, \cdots$, i. e. those curves of F on which lie the arcs $r_n q_{n+1}$, carried by k_n onto the line $y = +1$ in R_1, will cross the cross-section pq on γ at points r'_n which form a monotone sequence and which by our third requirement approach p. Also the points qq_1 determine an arc on a curve C_0, which maps under k_0 on $y = +1$. It follows that $S(\gamma)$, the set of all curves crossing γ, will contain all the sets U_n. Now for each n let a cross-section $r_n r_{n+1}$ be determined in U_n as the inverse image of the straight line in R_1 joining $k_n(r_n)$ and $k_n(r_{n+1})$. If we direct each curve C_n so that $\mathfrak{D}^{\#}(C_n)$ contains $\lambda(p)$ then $r_{n-1}r_n$ will lie in $\mathfrak{D}^{*}(C_n)$, except for the endpoint r_n which is on C_n. Hence the arcs $q_1 r_1$, $q_1 r_1 r_2$, $q_1 r_1 r_2 r_3, \cdots$ are each cross-sections by Theorem I 3.5-3, and they approach as limit an arc Γ from q_γ to infinity, which will be then a cross-section. The set U bounded by (1) $\lambda(p)$, (2) the cross-sectional arc pq on γ, (3) the arc qq_1 on C_0 and (4) the cross-section Γ will, as will be shown below, be an r-set interior to $V(\lambda(p))$, i. e. we shall exhibit the homeomorphism k of the definition.

(B) Now change the meaning of γ slightly to let γ denote only the arc pq on γ and, as above, let $S(\gamma)$ denote the set of curves of F crossing γ, and $S(\Gamma)$ the set of curves crossing Γ. We see from above that $S(\Gamma) \cup C^* = S(\gamma)$. Each of these sets is split into two domains if we remove γ, and we shall let $S'(\Gamma)$, $S'(\gamma)$ denote respectively the domains containing $\lambda(p)$. By Theorem 1. 1-1 there is a homeomorphism $k_1 \colon S'(\gamma) \to R_1''$, $R_1'' = \{(x, y) \mid -1 \leqq x < \infty, 0 \leqq y \leqq 1\}$ and $k_1(\gamma)$ is the line $x = -1$, $k_1(\lambda(p))$ is the x-axis for $x \geqq -1$. The curve C_0 on which q, q_1 lie will map onto the line $y = 1$ and Γ'', the image of Γ, will be an arc given by $x = \phi_2(y)$, $0 \leqq y < 1$, $k_1(q_1) = (\phi_2(1), 1)$ and $\lim\limits_{y \to 0} \phi_2(y) = \infty$.

Now we let $R_1' = \{(x, y) \mid -1 \leqq x < 1, 0 \leqq y \leqq 1\}$, i. e. a rectangle with the right side missing. We shrink $S'(\gamma)$ into R_1' along the lines $y = \text{constant}$ by the homeomorphism $k_2 \colon S'(\gamma) \to R_1'$ where $k_2 \colon (x, y) \to (x, y)$

for $-1 \leqq x \leqq 0$ and $k_2\colon (x,y) \rightarrow (x/(x+1), y)$ for $0 \leqq x < +\infty$. Let $x = \phi_1(y)$ be the function whose graph is Γ', the image of Γ'' under k_2. Γ' will be a half-open arc from the point $(\phi_1(1), 1)$ on the line $y = 1$ to the point $(1, 0)$ as limit point, i. e. as $y \rightarrow 0$ the point $(\phi_1(y), y)$ approaches $(1, 0)$ along Γ'. The side $x = -1$, the side $y = 0$, the segment from $(-1, 1)$ to $(\phi_1(1), 1)$ on $y = 1$, and finally Γ' will bound the image of U under k_2k_1. Denote this portion of R_1' by U'. Then we perform a final homeomorphism $k_3\colon U' \rightarrow \tilde{R}_1$ where $k_3\colon (x,y) \rightarrow \{(2+2x)/(1+\phi_1(y)) - 1, y\}$. The combined homeomorphism $k = k_3k_2k_1\colon U \rightarrow R_1$ is easily seen to be the desired homeomorphism as required in the definition of an extended r-set. Hence the proof of the theorem is complete. With these preliminaries completed we can prove the existence of the family G, complementary to F.

1. 2. Complementary curve families. Given a branched regular curve family F filling π, we shall call another such family, G, filling π *complementary* to F if (1) the singularities of G are exactly those of F and each is of the same type, i. e., a point b is an n-th order branch point of G if and only if it is an n-th order branch point of F; and (2) every curve of G is a cross-section of F. It follows at once from this definition and Theorem I 3. 2-4 that if G is complementary to F, then F is complementary to G. Hence we may speak of two complementary families, F and G, filling π. They will have a common set of singular points, B.

The major result of this section is to establish that every branched regular curve family F has a complementary family G. In [IV] it is shown that this is true when the set of singular points is empty, i. e. for a family F^* regular throughout a simply connected domain R^*, for we may by [IV] map F^* onto a family F' filling the xy-plane and defined by differential equations, $dx/dt = f(x,y)$; $dy/dt = g(x,y)$. The orthogonal trajectories define a family G' complementary to F' and the inverse image G^* of G' is then the desired complementary family to F^*. This result immediately gives us a family G^* complementary to F^* in $R^* = \pi - \tilde{\mathfrak{J}}$, $\tilde{\mathfrak{J}}$ being the cuts removed from π to make R simply connected [I]. The method we shall use to establish the existence of a family G complementary to F will be to consider first F^* and its complementary family G^*, both defined in R^* and then to modify G^* slightly near the boundary of R^*, i. e., near the cuts $\lambda(b)$, obtaining a family $\tilde{G}^*$ which becomes a family $\tilde{G}$ of the desired type when $\tilde{\mathfrak{J}}$, the set removed from π to give R^*, is replaced again. Theorem I 4. 1-3 tells us that we may cover $\tilde{\mathfrak{J}}$ with a collection $\{V[\lambda(b)]\}$ of disjoint open sets; we shall assume such a covering, to be fixed throughout what follows, and moreover, assume that each $V \subset U_\epsilon[\lambda(b)]$ an ϵ-neighborhood of $\lambda(b)$ where $\epsilon > 0$ is fixed. Any modification in G^* will actually take place deep inside V, i. e., in an open set whose closure lies in V. We shall actually discuss the modification for one such V and, assuming similar modifications have taken place in each V, we will denote by $\tilde{G}^*$ the modified G^*. $\tilde{G}^*$ will then be shown to be such that when $\tilde{\mathfrak{J}}$ is replaced, $\tilde{G}^*$ becomes a set $\tilde{G}$ complementary to F.

Restricting ourselves then to a definite $\lambda(b)$, and its neighborhood $V[\lambda(b)]$, as model, we proceed to define for $\lambda(b)$ a certain possibly infinite collection of closed sets $W_0, W_1, \cdots$ all contained inside $V[\lambda(b)]$ and surrounding $\lambda(b)$. These are the sets in which G^* will be modified. W_0 is an extended r-set, and if the number of curves in $\lambda(b)$ is finite, and only in this case, there will be a last set W_N of this collection which is also an extended r-set. All the other sets W_i will be r-sets in the sense of [I], which we shall hereafter call merely r-sets. These sets will be chosen as follows: *First*, let $b_0 = b, b_1, b_2, \cdots$ be the branch points on $\lambda(b)$, numbered so as to recede monotonely to infinity, and let the curves in R^* of each $St(b_i)$ be numbered with two indices, the first being that of b_i, the second being given by a counterclockwise numbering of the $St(b_i)$ proceeding from the first curve which follows counterclockwise after a curve of $St(b_i)$ not on C^* $(C^* \supset \lambda(b))$ to the last curve of $St(b_i)$ not on C^*: $C_{01}, \cdots, C_{0n_1}; C_{11}, C_{12}, \cdots, C_{1n_2}; \cdots$ etc.; let C_{00} denote C. (See Figure 2.) Then all numbered curves are in R^*. *Second*, choose regular points s_{ij} on each C_{ij} and short cross-sections γ_{ij} through s_{ij}, the γ_{ij} being in each case an arc on a curve of G^* and both s_{ij} and γ_{ij} being chosen so as to lie in $V(\lambda)$. Now we choose our sets W_n as follows: $W_0 \subset V(\lambda)$ is an extended r-set bounded on one side by an arc $r_{00}s_{00}$ on γ_{00} and on one side, of course, by $(s_{00}b) \cup \lambda(b)$. Next, in the domain bounded by the maximal chain determined by the adjacent curves C_{00}, C_{01} we choose an r-set $W_1 \subset V(\lambda)$ of the arc $s_{00}s_{01}$ on these curves, with W_1 bounded by the arcs $s_{00}t_{00}$ on γ_{00} and $r_{01}s_{01}$ on γ_{01}. Similarly, we choose $W_2, \cdots, W_{n_1}$, each an r-set contained in $V(\lambda)$ and bounded by arcs on two of the γ_{0i}'s. It may be that b_0 is the only branch point of $\lambda(b)$, in which case the next set W_{n_1+1} is the last and must be an extended r-set, bounded on one side by an arc $s_{0n_1}t_{0n_1}$ on γ_{0n_1}. Otherwise, we choose for W_{n_1} an r-set of $s_{0n_1}b_0b_1s_{11}$, an arc on the adjacent chain C_{1n_1}, C', C_{11}; C' being the curve of $\lambda(b)$ with endpoints b_0, b_1. The r-set W_{n_1} is so chosen that its cross-sectional ends are arcs $s_{0n_1}t_{0n_1}$ and $r_{11}s_{11}$ on γ_{0n_1} and γ_{11} respectively, and that it lies in $V(\lambda)$. This process is continued until we have chosen r-sets (or extended r-sets) on both sides of every curve of $St(b_i)$ for all b_i and hence, in particular, on both sides of each curve of $\lambda(b)$. Then $\lambda(b)$ will be contained in the interior of the set $W_\lambda = \bigcup_i W_i$. W_λ is bounded by an open arc Γ extending to infinity in each direction; and Γ consists in one case of *one* infinite cross-section of F^*, not in general a curve of G^*, plus an infinite number of arcs alternately on curves of F^* and on curves of G^* (the latter of the form $r_{ij}s_{ij}t_{ij} \subset \gamma_{ij}$); or else in

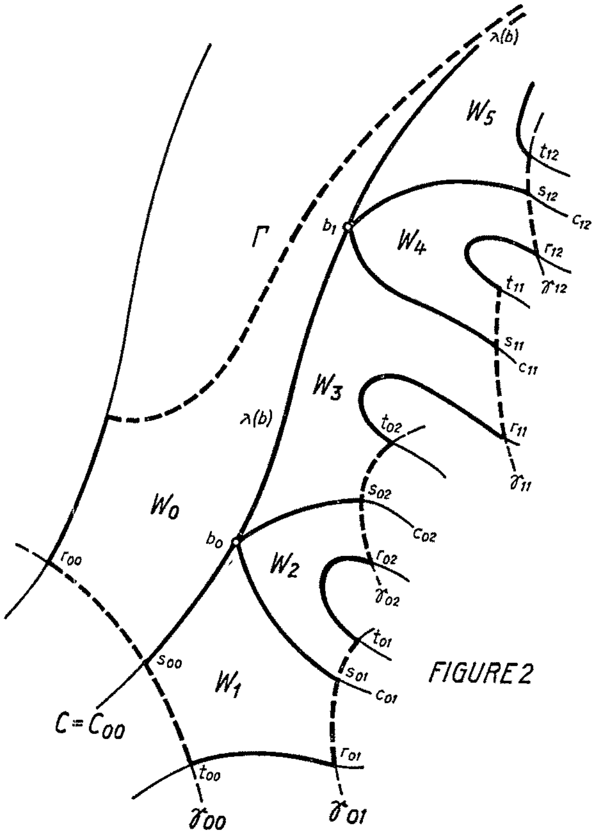


FIGURE 2

the other case, Γ consists of a finite number of such alternate arcs on F^* and G^* plus *two* half-open cross-sections of F^* extending to infinity. The first case occurs when there is one extended r-set and the number of sets W_i is infinite, the second when there are two extended r-sets and a finite collection of sets W_i comprising W_λ. Γ lies entirely inside $V(\lambda)$, and W_λ, which consists of Γ plus that one of its complementary domains inside $V(\lambda)$, is a closed set. The W_i's clearly intersect only on curves of F, namely on $\lambda(b)$ and on the arcs $b_i s_{ij}$ on each curve C_{ij} of $R^* \cap St(b_i)$ for b_i in $\lambda(b)$. We denote by $\tilde\lambda$ the set of all points which lie on the common boundary of two or more W_i's. A point of $\tilde\lambda$ which is a regular point clearly lies on the intersection of just two such sets, whereas each branch point b_i lies on the intersection of $2m$, where m is the multiplicity of b_i. We denote by $W_i{}^*$ the set $W_i - \tilde\lambda$, and by $W_\lambda{}^*$ the set $W_\lambda - \tilde\lambda$, and finally by $V_\lambda{}^*$ the set $V(\lambda) - \tilde\lambda$. Then let $\bar G^* = G^*[V_\lambda{}^*]$ and $\bar F^* = F^*[V_\lambda{}^*]$, i. e. we remove from $V(\lambda)$ all points on two or more sets W_i.

Now each W_n has associated with it a homeomorphism k_n, of W_n onto R_1 or, if it is an extended r-set, onto $\tilde R_1$. In order that the modification of G^* to $\hat G^*$ which we are going to make will not destroy the relationship between G^* and F^* we will actually achieve it by a homeomorphism h of $\tilde R^*$ ($\tilde R^* = R^* - \bigcup\limits_{\lambda \subset \tilde T} \tilde\lambda$) onto itself, which is the identity outside of each set W, but which inside such a set carries each curve of $\bar F^*$ onto itself. The need for this modification arises from the fact that although $\bar F^*$, $\bar G^*$ are complementary in $\tilde R^*$, and hence in $V^*(\lambda) = V(\lambda) - \tilde\lambda$, they will not in general be complementary in $V(\lambda)$ along points of $\tilde\lambda$. In fact replacing $\tilde\lambda$ in V^* will not in general transform $\bar G^*$ into a regular curve family in $V(\lambda)$, since after all those points of $\tilde\lambda$ on λ are boundary points of the region R^* filled by G^*, and curves of G^* may have common endpoints on the boundary of the domain of G^*, or no endpoints (i. e. may extend to ∞ which is also a boundary point of R^*). Our procedure is to cut the plane along each $\tilde\lambda$, cutting along curves of F, i. e. whenever we cut a curve of G^*, we cut across it: in particular in cutting along $\tilde\lambda - \lambda$, since this is in R^*, we cut across curves of G^*. Then keeping the curves of F fixed (not pointwise) we move the "cut ends" of curves of G^*, with their individual points "sliding" along curves of F, into such positions that each regular point on $\tilde\lambda(b)$ becomes the endpoint of exactly one curve of G^* from each side of $\tilde\lambda$, and the branch points of multiplicity m the endpoint of $2m$ curves, one from each sector. Then replacing $\tilde\lambda$, $\bar G^*$ the modified G^* becomes a regular family at every regular point of F and is in fact complementary to F. We shall describe this operation piecewise, for each $W_n{}^*$ and, in fact, at first as a homeomorphism on the image of $W_n{}^*$ in R_1 (or $\tilde R_1$ as the case may be), (I) for r-sets, (II) for extended r-sets.

(I) We begin by defining a typical homeomorphism f_I in R_1 on the image under k_i of $\bar F^*[W_i{}^*]$, $\bar G^*[W_i{}^*]$, W_i an r-set. The image of $W_i{}^*$ will be $R_1{}^* = R_1 -(x\text{-axis})$, and we will denote the images of the curve families as $F_1{}^*$, $G_1{}^*$, respectively. The former will, of course, be just the lines $y = a$, $0 < a \leqq 1$, and the latter will be a regular curve family filling $R_1{}^*$, complementary to $F_1{}^*$, and having among its curves the two lines $x = \pm 1$, images of arcs γ_{ij}, which lie on curves of G^*. It will be seen that $G_1{}^*$ consists exactly of the curves whose inverse images cross C', the inverse image of $y = 1$ in $R_1{}^*$, for, if we consider any curve of $\tilde{G}^*$ with a point inside W_i, it is clear that it must leave W_i in each direction, there being no boundary points of $\tilde{R}^*$ interior to W_i; and hence, it must either cross C' or have two endpoints on $\tilde{\lambda}(b)$. It could scarcely have both endpoints on $\tilde{\lambda}(b)$, however, without crossing some curve of F^* twice inside W_i, which is impossible since the curves of G^* are cross-sections of F^*. Moreover, no curve of G^* will cross C' more than once, since C' is a cross-section of G^*. Thus we may define a function f_I mapping $R_1{}^*$ onto itself as follows: Let $\bar{x} = f(x, y)$ be defined by $f(x, 1) \equiv x$ and $f(x, y) =$ constant on each curve of $G_1{}^*$, and let $\bar{y} = g(x, y)$ be defined by $g(x, y) \equiv y$. Then it follows from the above remarks and the work of Kaplan [II] and [III] that $f_I \colon (x, y) \rightarrow (\bar{x}, \bar{y})$ is a homeomorphism of $R_1{}^*$ onto itself which takes each curve of $F_1{}^*$ onto itself, and each curve of $G_1{}^*$ onto a line $x = b$, $-1 \leqq b \leqq 1$, the lines $x = \pm 1$ being held pointwise fixed, as is the line $y = 1$, i. e., all of the boundary of $R_1{}^*$ on which f_I is defined is held pointwise fixed. $h \,|\, W_i{}^*$ is then defined by $k_i{}^{-1} f_I k_i$, and if thus defined h maps $\tilde{F}^*[W_i{}^*]$ onto itself, takes $\tilde{G}^*[W_i{}^*]$ homeomorphically onto a new family $\tilde{G}^*[W_i{}^*]$ which is still complementary to F^* and which is identical to $\tilde{G}^*$ on the boundary of $W_i{}^*$. Since k_i is actually a homeomorphism of all of W_i onto R_1, it will now map $F[W_i]$ and $\tilde{G}^*[W_i]$ so that the curves $F^*[W_i]$, $\tilde{G}^*[W_i]$ will map onto the lines $y = a$ and $x = b$, respectively. We re-denote k_i by $\tilde{k}_i$ to emphasize that it acts on $\tilde{G}^*$. Thus it is clear that every curve of $\tilde{G}^*[W_i]$ has exactly one endpoint, unique to it, on $\tilde{\lambda}$ and exactly one endpoint unique to it on the curve of F^* forming the opposite side of W_i. The regularity of $\tilde{G}^*$ which we have achieved at $\tilde{\lambda}$ is precisely what is needed. We assume a similar homeomorphism defined for every index i such that W_i is an r-set; then h will be defined on every set of W except the one (or two) extended r-set(s).

(II) Now let us suppose that we are dealing with an extended r-set say W_0, with its associated homeomorphism k_0 onto $\tilde{R}_1$. Again let $F_1{}^*$, $G_1{}^*$ denote the images of the respective families of W_0 in $\tilde{R}_1{}^* = \tilde{R}_1 -(x\text{-axis})$, $F_1{}^*$ being just the lines $y = a$; the line $x = -1$ in $\tilde{R}_1{}^*$, but not in general the line $x = +1$, being a curve of $G_1{}^*$. f_{II} will be given as the composition of four homeomorphisms of $\tilde{R}_1{}^*$ onto itself. Before we can describe f_1, the first of these, we must note that there is in W_0 at least one curve ψ of G^*, distinct from the arc $r_{00}s_{00}$ on γ_{00} (the inverse image of $x = -1$), whose image ψ_1 in $\tilde{R}_1$ joins a point $(x'', 0)$ to a point $(x', 1)$, where $-1 < x'', x' < 0$,

i. e., a curve of G^* joining one side of W_0 to the other, and intersecting each side at a regular point of R^*, in particular, on C_{00}, not on $\lambda(b)$ (see Figure 3). That such a curve exists follows from the fact that in the family G^*, regular in R^*, the arc $r_{00}s_{00}$ on a curve of G^* has an r-neighborhood U (by Theorem I 1. 2-2) with $\bar{U} \subset R^*$. The curves $C_{s_{00}}$ and $C_{r_{00}}$ (see Figure 2) have small arcs entirely in this neighborhod, since they are cross-sections of G^*, and each of these will be crossed by an infinite number of curves of G^* on each side of $s_{00}t_{00}$, one of which will serve our purpose; namely, one crossing for each of these arcs that part which is the inverse image respectively of the segments $(-1, 1)$ to $(0, 1)$ and $(-1, 0)$ to $(-\epsilon, 0)$, $1 > \epsilon > 0$. ψ_1 will be given by a continuous function $x = \psi_1(y)$, $0 \leqq y \leqq 1$, and we shall use it to define $f_1 : R_1^* \to R_1^*$ given by f_1: $(x, y) \to (\bar{x}, \bar{y})$ where

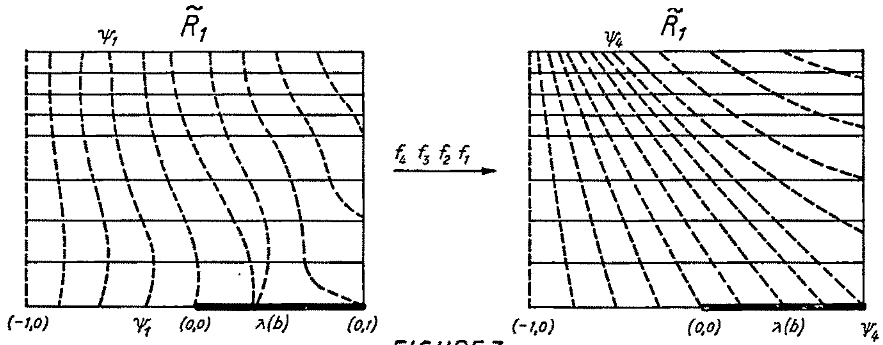


FIGURE 3

$$\bar{x} = \{[1 + \psi_2(y)]x - [\psi_1(y) - \psi_2(y)]\}/[1 + \psi_1(y)] \text{ for } -1 \leqq x \leqq \psi_1(y),$$

$$\bar{x} = \{[1 - \psi_2(y)]x - [\psi_1(y) - \psi_2(y)]\}/[1 - \psi_1(y)] \text{ for } \psi_1(y) \leqq x \leqq +1;$$

$$\bar{y} \equiv y$$

(where $\psi_2(y) = (x' - x'')y + x''$, this being the equation of the line joining $(x', 1)$ to $(x'', 0)$, the curve into which ψ_1 is mapped by f_1).

The next homeomorphism, $f_2 : R_1^* \to R_1^*$ will carry ψ_2 into ψ_3, the line $x = x'$. f_2 is given by f_2: $(x, y) \to (\bar{x}, \bar{y})$ where:

$$\bar{x} = \{(1 + x')x + [x' - \psi_2(y)]\}/[1 + \psi_2(y)] \text{ for } -1 \leqq x \leqq \psi_2(y),$$

$$\bar{x} = \{(1 - x')x + [x' - \psi_2(y)]\}/[1 - \psi_2(y)] \text{ for } \psi_2(y) \leqq x \leqq 1;$$

$$\bar{y} \equiv y.$$

Each of these homeomorphisms holds the boundary curves $x = \pm 1$, $y = 1$ pointwise fixed. To describe f_3 we first denote by M that portion of $\tilde{R}_1{}^*$ which lies on or to the left of ψ_3, i. e., $M = \{(x, y) \mid -1 \leqq x \leqq x', 0 \leqq y \leqq 1\}$. M is bounded on each side by a line $x =$ constant, i. e. $x = -1$, and $x = x'$, each the image of a curve of G^* under the composition of the above maps, and M is bounded on top and bottom by images of curves of F^*. The image of F^* in M is the family of lines $y = a$. Hence by precisely the same argument as in the definition of f_I for the neighborhood of type I above, we may find a homeomorphism $f_3 : M \to M$ which holds the boundary of M pointwise fixed, takes each curve $y = a$ onto itself, and takes the image family of G^* onto the lines $x = b$, $-1 \leqq b \leqq x'$. We extend f_3 to all of $R_1{}^*$ by defining it as the identity on the rest of this set. Again, f_3 will be a homeomorphism leaving the boundary curves $x = \pm 1$, $y = 1$, as well as the curve ψ_3 and all of $\tilde{R}_1{}^*$ to the right of ψ_3, pointwise fixed.

Finally, we define a homeomorphism $f_4 : \tilde{R}_1{}^* \to \tilde{R}_1{}^*$, again by giving $f_4 : (x, y) \to (\bar{x}, \bar{y})$ as follows:

$$\bar{x} = (\psi_4(y) + 1)((x + 1)/x' + 1)) - 1 \text{ for } -1 \leqq x \leqq x',$$

$$\bar{x} = (1 - \psi_4(y))((x - x')/(1 - x')) + \psi_4 \text{ for } x' \leqq x \leqq +1;$$

$$\bar{y} = y,$$

where ψ_4 denotes the line $x = \psi_4(y) = (x' - 1)y + 1$ joining $(x', 1)$ to $(1, 0)$, this being the image of ψ_3 under f_4. The image of M under f_4 will be denoted by M_1 and will be the trapezoid bounded by ψ_4, the x-axis, the line $x = -1$, and the segment from $(-1, 1)$ to $(x', 1)$ on the line $y = 1$. f_4 takes the lines $y = a$ onto themselves, and the lines $x = b$, $-1 \leqq b \leqq x'$ of M onto a family of nonintersecting straight lines joining the points of the top edge of M_1 to the bottom (as listed above). f_4 leaves the lines $x = \pm 1$ and $y = 1$ pointwise fixed.

Now we define $f_{II} : \tilde{R}_1{}^* \to \tilde{R}_1{}^*$ as the homeomorphism $f_4 f_3 f_2 f_1$, and we define $h \mid W_0{}^*$ as $k_0{}^{-1} f_{II} k_0$ (see Figure 3). Then $h \mid W_0{}^*$ is a homeomorphism of $W_0{}^* = W_0 - \tilde{\lambda}$ onto itself which is pointwise fixed on the boundary of $W_0{}^*$ in $\tilde{R}^*$, i. e., on $t_{00} s_{00}$, on $C_{t_{00}}$, and on the extended cross-section which bounds one side of W_0. h also takes the curves of $G^*[W_0{}^*]$ homeomorphically onto a family $\tilde{G}^*$, at the same time mapping each curve of F^* onto itself. Now, if as above for k_i, we re-denote k_0 by $\tilde{k}_0$, then we have a homeomorphism of all of W_0 onto $\tilde{R}_1$ which takes $\tilde{\lambda}$ onto the x-axis between $(-1, 0)$ and $(1, 0)$, with b_0 mapping onto $(0, 0)$, and s_{00} onto $(-1, 0)$, and which moreover, takes the curves of F onto the lines $y = a$ and takes part of $\tilde{G}^*$ onto the straight lines joining the top and bottom of M_1 as described above, the remainder of $\tilde{G}^*$ mapping onto a regular family filling the rest of R_1. The curve Ψ of $\tilde{G}^*$, image of ψ under $h \mid W_0^*$ divides W_0 into two domains, one of which maps onto M_1, the other onto $R_1 - M_1$. We shall denote the one which maps onto M_1, together with its boundary, by $\tilde{W}_0$, the boundary consisting of two curves of $\tilde{G}^*$, namely $r_{00}s_{00}$ and Ψ, together with $C_{r_{00}}$ and $s_{00}b_0 \cup \lambda(b)$ in F. It is obvious that $M_1 \subset \tilde{R}_1$ can be mapped onto $\tilde{R}_1$ by a homeomorphism g which holds $x = -1$ and $y = 0$ pointwise fixed, takes each line $y = a$ into itself, and finally moves the image curves of $\tilde{G}^*$ in M_1 onto the lines $x = b$, $-1 \leqq b \leqq 1$, keeping, of course, their lower endpoints fixed, thus taking the line Ψ onto $x = 1$. Then $g\tilde{k}_0 \colon \tilde{W}_0 \rightarrow \tilde{R}_1$ with F going onto the lines $y =$ constant and $\tilde{G}^*$ onto the lines $x =$ constant. W_0 is then again, like W_0, an extended r-set of $\lambda(b)$, but of a kind which is bounded by curves of two complementary families and has associated a homeomorphism $g\tilde{k}_0$ which maps the curves of the respective families onto the lines parallel to the axes in $\tilde{R}_1$. Hereafter, we shall denote $g\tilde{k}_0$ merely by $\tilde{k}_0$. Now if W_N is a second extended r-set in W, then it must be the last W_i defined for $\lambda(b)$ and on it we define, in a manner entirely parallel to the above discussion, f_{II}, $h \mid W_N^*$, $\tilde{W}_N$, $\tilde{k}_N$, etc.

Thus we have defined $h \mid W_i^*$ for all i. Now since the W_i^* are overlapping closed sets of V_λ^* with only a finite number of the sets W_i containing any given point, and since h is actually the identity along their overlapping boundaries as well as on Γ, the boundary of W_λ, we have defined a homeomorphism h of $W_\lambda^* = W_\lambda - \tilde{\lambda}$ onto itself. Assume that h is similarly defined for a set $W_\lambda^* \subset V[\lambda(b)]$ for *every* cut $\lambda(b)$ contained in $\tilde{\mathfrak{T}}$, and define h as the identity outside the W_λ's. We remark that the collection of all the sets W_λ for $\lambda(b)$ in $\tilde{\mathfrak{T}}$, together with the set $\pi - \bigcup\limits_{\lambda \varepsilon \tilde{T}} W_\lambda$, is a collection of overlapping closed sets which has a locally finite character, i. e., every neighborhood of any point meets only a finite number of the closed sets. This is clear because the cuts, λ, recede to infinity, and each W_λ lies in an ϵ-neighborhood of the cut λ, $\varepsilon > 0$ being fixed. Then it follows that h is a homeomorphism of $\tilde{R}^*$ onto itself, where by $\tilde{R}^*$ we mean $R^* - [\bigcup\limits_{\lambda \varepsilon \tilde{T}} \tilde{\lambda}(b)]$. h carries every curve of F^* onto itself homeomorphically, and every curve of $G^*[R^*]$ homeomorphically onto a family $\tilde{G}^*$ which is complementary to F^* in $\tilde{R}^*$ and which coincides with G^* except in the interior of the W_λ's.

It remains to prove that by adding the boundary points of $\tilde{R}^*$, i. e., $\bigcup\limits_{\lambda \varepsilon \tilde{T}} \tilde{\lambda}$, the curves of $\tilde{G}^*$ become curves of a family $\tilde{G}$ complementary to F in π. To prove this we must first prove that $\tilde{G}$ is regular in $R = \pi - B$. Now if v is a point of $\tilde{R}^*$, this is clear, since $\tilde{G} = \tilde{G}^*$ (which is homeomorphic to $\tilde{G}^*$)

in some neighborhod of p. In fact, it is clear (from the method used in getting the homeomorphism f_I above) that there is an arbitrarily small r-neighborhod of p whose closure maps onto $R_0 = \{(x, y) \mid |x| \leqq 1, |y| \leqq 1\}$ so that the lines $x =$ constant are the images of the curves of $\tilde{G}$, those lines $y =$ constant are the image curves of F.

Now, however, suppose that p is a regular point on $\bar{\lambda}(b)$. Then p will be on the common boundary of just two of the neighborhoods W_i, since p is not a branch point. Let W_n, W_m be the two neighborhoods. Then p is interior to $W_n \cup W_m$, and it follows from Theorem I 1. 2-3 that $\tilde{G}[W_n \cup W_m]$ is regular at p, since $\tilde{G}$ is regular in W_n and in W_m separately, as may be seen from the existence of the maps $\tilde{k}_n, \tilde{k}_m$ onto R_1 (or $\tilde{R}_1$ as the case may be) with $\tilde{G}$ mapping onto the lines $x =$ constant. Thus $\tilde{G}$ is regular at every point of R, so that the singularities of $\tilde{G}$ are contained in the set B of singularities of F, and are thus isolated. Now each branch point is in a cut, and hence will be $b_i \varepsilon \lambda(b)$ for some i and some $\lambda(b)$. b_i is on the common boundary of just $2m$ sets W_n, where m is the multiplicity of b_i. Then it is clear that there are just exactly $2m$ curves of $\tilde{G}[W]$, exactly one in each of these $2m$ sets, which have b_i as a limit point in one direction. For, if W_n has b_i on its boundary, then in the homeomorphism $\tilde{k}_n \colon W_n \rightarrow R_1$ the point b_i will map onto a point $(a, 0)$ and the inverse image of the line $x = a$ is the single curve of $\tilde{G}[W_n]$ which has b_i as a limit point. It follows at once that b_i is a branch of multipilicity $2m$ of $\tilde{G}$. Hence we have established that $\tilde{G}$ is a branched regular curve family with the same branch points as F. Again, just as above, it is clear that it is possible to find an arbitrarily small neighborhood U of each b_i which is homeomorphic to $|z| < 1$, and moreover, with a homeomorphism k carrying $F[\bar{U}]$ onto the level curves of $\Re(z^m)$ and $\tilde{G}[\bar{U}]$ onto the level curves of $\Im(z^m)$.

Finally, to complete the proof that $\tilde{G}$ is complementary to F, we note that by Corollary 2 to Theorem I 3. 5-3 we have at once that every curve of $\tilde{G}$ is a cross-section of F. This completes the proof of the following:

THEOREM 1. 2-1. *Every branched regular curve family* F *has at least one complementary family* G $(= \tilde{G})$ *as described above.*

1. 3. The fundamental theorem. Given any branched regular curve family F on π, we have shown the existence of a complementary family G; and also, we have shown [I] that each of these families is the level curve family of a continuous function (without relative extrema) $f(p)$ and $g(p)$ respectively. This enables us to define a single-valued mapping T_1 from the plane π to the complex w-plane as follows: $T_1(p) = u + iv$ where $u = f(p)$

and $v = g(p)$. $T_1(p)$ is clearly continuous, because f and g are continuous. Moreover, T_1 is locally a homeomorphism on R and is exactly m-to-1 in the neighborhood of an m-th-order branch point. To show this, it is sufficient to consider the special neighborhoods mentioned in the proof of the previous theorem, i. e., for every regular point we consider only a neighborhood U such that there is a homeomorphism of U onto the rectangle R_1 of the xy-plane such that $F[U]$ goes onto the lines $y =$ constant and $G[U]$ onto the lines $x =$ constant. Then T_1 becomes a map of R_1 onto a rectangle in the uv-plane carrying the lines $y =$ constant onto $u =$ constant and $x =$ constant onto $v =$ constant. It is clearly a homeomorphism since it is monotone on each line $x =$ constant and each line $y =$ constant. This is exactly as in [III]. It is equally easy to show that in a neighborhood V of a branch point, where $F[V]$ and $G[V]$ map onto $\Re(z^m)$ and $\Im(z^m)$ respectively under a homeomorphism of V onto $|z| < 1$, T_1 carries V onto an open set and is at most m-to-1, where m is the multiplicity of the branch point (cf. [III]). Hence T_1 is not only interior but light (since for every point there is a neighborhood in which f and g take on the same value only a finite number of times in the neighborhood). It follows from Stoilow [V, Chapter V, part III, §5] and Whyburn [VII] that T_1 is topologically equivalent to an analytic function $W = \phi(z)$, i. e., there exists a homeomorphism $p = h(z)$ of the plane π onto either the domain $D_1 = \{z| \ |z| < 1\}$ or $D_\infty = \{z| \ |z| < \infty\}$ of the z-plane such that $\phi(z) = T_1[h(z)]$ is analytic. The family F' of level curves of the real part of $\phi(z)$ are just those curves mapping onto the lines $u =$ constant of the w-plane and hence are homeomorphic to F under h. It is thus proved that:

THEOREM 1. 3-1. *Given any branched regular curve family F there exists a function harmonic in either the finite plane or the unit circle whose level curves are homeomorphic to F.*

Since, if the function $u(x,y)$ is harmonic in a domain D, it is differentiable in D, its level curves will satisfy the differential equations $dx/dt = u_y$, $dy/dt = -u_x$, we have at once:

THEOREM 1. 3-2. *Given any branched regular curve family F, then there is a solution family of a system of differential equations to which it is homeomorphic.*

2. Decomposition of F into Half-Parallel Subfamilies.

2. 1. Extended cross-sections.

THEOREM 2. 1-1. *Let p be any regular point of π, C_p the curve of F through p, and let C be a curve containing a point q such that there is a cross-section pq. Then there will be a cross-section from p to an arbitrary point q' of T_C if and only if $q' \varepsilon C^*$, where C is directed so that $p \varepsilon \mathfrak{D}^*(C)$. Moreover, if $q' \varepsilon C^*$ and $U(qq')$ is any r-set (in $\mathfrak{D}^*(C)$) of qq', we may choose the cross-section qq' as follows: $qq' \equiv qrq'$ where qr lies on pq and rq' is in $U(qq')$.*

Proof. Suppose q' to lie on C^* and let $U(qq')$ be any r-set of qq'. Now moving along pq from p, the cross-section pq lies entirely inside $U(qq')$ from some point on, so we may choose some r on pq, with rq interior to U, letting prq now denote pq. We direct C_r so that $pr \subset \mathfrak{D}^*(C_r)$ and $rq \subset \mathfrak{D}^\#(C_r)$, which we can do by Theorem I 3. 5-3 since prq is a cross-section. We replace rq by a cross-section rq' in U which is found as follows: U by definition is homeomorphic to the rectangle R_1 in the xy-plane, and we join in R_1 the image of r to that of q' by a straight line, whose inverse image we then take for rq'. Since the straight line is a cross-section of the image of F, i. e., the lines $y = k$, rq' will be also a cross-section, and will lie in the same domain $\mathfrak{D}^\#(C_r)$ as rq, since each cross the same curves in U. Hence, by Theorem I 3. 5-3, we know that prq' is a cross-section.

It remains only to prove that if C' is any curve of T_C not on C^*, then there is no cross-section to C' from p. Now p lies in $\mathfrak{D}^*(C^*)$ and C' in $\mathfrak{D}^\#(C^*)$, hence any such cross-section, if it existed, would have to cross C^* and thus would have two points on T_C, contrary to the assumption that it is a cross-section.

THEOREM 2. 1-2. *Let the trees of F be numbered as in* [I, *Section* 4], *i. e., in a standard numbering, using the concentric circles K_n of center p and radius n; further, let the cuts $\mathfrak{J}$ be removed from F, leaving $F^* = F[R^*]$. Then, outside every circle K_n lies at least one curve of F^* which can be reached from p by a cross-section lying in $R^* \cap \mathfrak{D}^*(C_p)$.*

Proof. Denote by $\{C\}$ the collection of all curves in $\mathfrak{D}^*(C_p)$ which can be reached by a cross-section from p lying in $R^* \cap \mathfrak{D}^*(C_p)$. We direct each curve of $\{C\}$ so that $\mathfrak{D}^*(C) \subset \mathfrak{D}^*(C_p)$. The existence of a cross-section from p to $q \varepsilon C$ makes this possible, i. e., direct C so that $\mathfrak{D}^\#(C) \supset pq$. $\{C\}$ will certainly not be empty since we assume p to be a regular point.

Now define on the curves of $\{C\}$ the positive real-valued function $d(C) = \underset{x \varepsilon C}{\text{g. l. b.}}$ {distance from x to p}. We have at once that C is outside K_n if and only if $d(C) > n$. Also it is clear that $\mathfrak{D}^*(C) \supset \mathfrak{D}^*(C')$ implies that $d(C) < d(C')$. To prove the theorem we must show that the numbers $d(C)$ are unbounded. We asume that this is not so; then there is a least upper bound d' of $d(C)$ for C in $\{C\}$. To show that this is impossible we shall choose $N > d'$ and consider intersections of curves of $\{C\}$ with K_N. Since we have assumed $d(C) \leqq d' < N$, every curve of $\{C\}$ will intersect K_N, although by Theorem I 4. 1-1 only a finite number of these curves lie completely inside K_N. All but a finite number of curves of $\{C\}$ in fact, not only have both endpoints outside K_N, but are themselves the only curve of T_C intersecting K_N. Hence, we may choose an infinite sequence of curves C_m of $\{C\}$ such that $d(C_m) \to d'$, $T_{C_m} \cap K_N \equiv C_m \cap K_N$, and $C_m \cap K_N$ contains neither endpoint of C_m. Having chosen such a sequence we find a subsequence q'_n of points $q'_n \varepsilon C_n$ which approach a regular point q as a limit and all lie on one side of the image of C_q in an r-neighborhood $U(q)$ (i. e., in the upper or lower half of R_0, the image of $U(q)$). This may be done as follows: first, by compactness of K_N we may find $q_{m_i} \varepsilon C_{m_i} \cap K_N$ (a subsequence of the m's) which converges to some point q'. Second, if q' is a regular point, we let $q = q'$ and choose a subsequence q'_n of the q_{m_i}'s, all of whose points lie in one side only of $U(q)$. Or, if q' is a branch point, let $V(q')$ be any admissible neighborhood of q'; then an infinite subsequence of the q_{m_i}'s will lie in one sector of V. If q is any regular point on either of the adjacent curves bounding this sector of V, there will be a corresponding sequence of points q'_n on the same curves C_n which contain the points q_n and such that $q'_n \to q$. The q'_n will lie on the same side of C_q in any r-neighborhood of q and is thus the desired sequence. Finally, we may choose a subsequence of q'_n which we will denote by r_n such that if qs is a cross-section from q to s in $U(q)$, where s lies on the same side of $U(q)$ as the q'_n, then the intersections $C_n \cap qs$ tend monotonely to q on qs (C_n denoting the curve on which r_n lies). Thus we have $d(C_n) \to d'$ monotonely since $\mathfrak{D}^*(C_n) \supset \mathfrak{D}^*(C_{n+1}) \supset C_q$ for all n. We direct C_q so that $\mathfrak{D}^*(C_n) \supset \mathfrak{D}^*(C_q)$.

Now choose in $\mathfrak{D}^\#(C_q)$ an r-set W of qq'' where q'' is any point of $C^\#_q$ which is in R^*. W is chosen so that its interior lies in R^*, which is possible by Theorem I 4. 1-4. Now for $n \geqq n_0$, r_n will lie in W, and since we have $\mathfrak{D}^*(C_{n_0}) \supset C^\#_q$ and $\mathfrak{D}^\#(C_{n_0}) \supset C_p$, we may extend the cross-section $pr_{n_0} \subset R^* \cap \mathfrak{D}^*(C_p)$ to a cross-section $pr_{n_0}q'' \subset R^* \cap \mathfrak{D}^*(C_p)$ by merely adding to it that cross-section $r_{n_0}q''$ in $W \cap \mathfrak{D}^*(C_{n_0})$ which is the inverse image in W of the straight line joining the images of r_{n_0} and q'' in R_1, the image of W. This will be a cross-section by Theorem I 3. 5-3. Now since q'' is a regular point of a curve $C_{q''}$, if we take its direction such that $C^{\#}_{q''} \equiv C^{\#}_{q}$, we have $\mathfrak{D}^{\#}(C_{q''}) \supset C_p, C_n$; and $\mathfrak{D}^*(C_{q''}) \subset \mathfrak{D}^*(C_n)$ for all n, whence $d(C_{q''}) \geqq d'$. Now it is easy, however, by taking an r-neighborhood of q'' (which will lie in R^*) to extend $pr_{n_0}q''$ to a slightly larger cross-section $pr_{n_0}q''s$, and since $C_s \subset \mathfrak{D}^*(C_{q''})$, we have at once that $\mathfrak{D}^*(C_{q''}) \supset \mathfrak{D}^*(C_s)$, where C_s is directed as a curve of $\{C\}$, i. e. so that $\mathfrak{D}^*(C_s) \subset \mathfrak{D}^*(C_r)$. Hence $d(C_s) > d(C_{q''}) \geqq d'$. This is contrary to the assumption that d' is a bound of $d(C)$. Hence $d(C)$ is unbounded, which is what was to be proved.

By an *extended cross-section,* we shall mean any curve in $R = \pi - B$ which meets each curve of F at most once and tends to infinity in one or both directions. An extended cross-section is said to *tend properly to infinity* in R in a given direction on it, if it tends to infinity in that direction in such a way that the curves meeting it tend uniformly to infinity with their intersection points with the cross-section. We shall also speak of an *extended cross-section* in R^* which will be an extended cross-section as above, and lie entirely in $R^* = \pi - \tilde{\mathfrak{J}}$, i. e., it meets only curves of F^*.

THEOREM 2. 1-3. *If p is any regular point on a curve C of F^*, then there is an extended cross-section in R^* from p, which lies in $\mathfrak{D}^*(C_p)$ and tends properly to infinity.*

Proof. We consider a curve C in F^* and p any point on it. As before K_n will denote a circle with center at p and radius n; and for any point s we shall let $Q_n(s)$ denote a circle with center at s and radius so chosen that $Q_n(s)$ contains K_n. Now we choose a regular curve C_1 in $\mathfrak{D}^*(C_p) \cap R^*$ for which there is a cross-section pq_1 in $\mathfrak{D}^*(C_p) \cap R^*$ from p to q_1 on C_1. Direct C_1 so that $\mathfrak{D}^*(C_p) \supset \mathfrak{D}^*(C_1)$ and choose in $\mathfrak{D}^*(C_1) \cap R^*$ a curve C_2 outside of $Q_1(q_1)$ and such that a cross-section q_1q_2 in $\mathfrak{D}^*(C_1) \cap R^*$ exists with q_2 on C_2. Having chosen C_n and $q_n \varepsilon C_n$ in this manner, we choose for C_{n+1} any regular curve outside of $Q_n(q_n)$ for which there is a cross-section q_nq_{n+1} in $\mathfrak{D}^*(C_n) \cap R^*$ to q_{n+1} on C_{n+1}. We direct C_{n+1} so that $\mathfrak{D}^*(C_n) \supset \mathfrak{D}^*(C_{n+1})$. We continue this process indefinitely by Theorem 2. 1-2. Then the curves $pq_1, pq_1q_2, pq_1q_2q_3 \cdots$ will all be cross-sections by Theorem I 3. 4-5. They approach a curve Γ extending from p to infinity in $\mathfrak{D}^*(C_p) \cap R^*$ which is an extended cross-section extending from p to infinity in R^*. The curves intersecting Γ tend uniformly to infinity with any sequence of their points of intersection tending to infinity on Γ; since if r on Γ is beyond q_n, then C_r lies outside K_n. Thus Γ is an extended cross-section tending properly to infinity in R^*.

2. 2. Half-parallel subfamilies of F. We mean by a *half-parallel subfamily of F* the collection of all curves of F which intersect an extended cross-section Γ tending from a point p on a curve C_p properly to infinity. And we shall mean by a *complete* half-parallel subfamily of F the curve $C_p{}^*$ together with all curves of F crossing Γ (C_p being so directed that $\mathfrak{D}^*(C_p) \supset \Gamma$). The first of these sets is homeomorphic to the lines $y = k$, $k \geqq 0$ of the half-plane by Theorem 1. 1-1 and the same reasoning as used in the proof of that theorem will establish this homeomorphism for the second case, the complete half-parallel subfamily also. The first will be denoted by S and the second by S^*. Clearly $S^* \supset S$ and when C_p is a regular curve they are identical. C_p is called the *initial curve* of S, $C_p{}^*$ the initial curve of S^*.

If $\Gamma(q)$ is any half-open cross-section of F tending from a regular point q properly to infinity, then the boundary of $S(\Gamma)$, $S(\Gamma)$ being the collection of curves intersecting Γ, is best described in terms of maximal chains C^*, $C^\#$ and the sets $\delta(C+)$, $\delta(C-)$ defined in [I, Section 3]. We shall refer to these latter two sets as *mixed maximal chains,* since they consist of two subchains of maximal chains, one clockwise adjacent, the other counterclockwise adjacent, e. g., $\delta(C+) = \delta^*(C+) \cup \delta^\#(C+)$ (which may be empty). $\delta(C)$ will denote $\delta(C+) \cup \delta(C-)$; it is empty if and only if C is a regular curve.

THEOREM 2. 2-1. *The boundary of $S(\Gamma)$ is a collection of maximal chains C^*, $C^\#$ and mixed maximal chains $\delta(C)$, where $\delta(C)$ is on the boundary if and only if C is in $S(\Gamma)$. From each set T_C of F there is either* (1) *no point,* (2) *exactly one maximal chain, or* (3) *a set $\delta(C)$ of T_C on the boundary of $S(\Gamma)$.* (1), (2) *and* (3) *are mutually exclusive.*

Proof. Suppose $C \varepsilon S(\Gamma)$ is a singular curve, then $\delta(C)$ is in the boundary of $S(\Gamma)$, for if we consider any point q on $\delta(C)$ there exists an r-set $U(pq)$ containing q and $p = C \cap \Gamma$ (since C lies on an adjacent chain with C_q); choosing a sequence of points $p_n \to p$, $p_n \varepsilon U \cap \Gamma$, we can find by Theorem I 3. 5-2 a sequence $q_n \varepsilon U$ such that $q_n \varepsilon C_{p_n}$ for all n and $q_n \to q$. Whence q is a limit point of points of $S(\Gamma)$. But, if q is in $\delta(C)$, it is on a curve of T_C other than C; and C_q therefore cannot intersect Γ. Hence C_q is not in $S(\Gamma)$, and thus q is on the boundary of $S(\Gamma)$. Moreover, no other curves of T_C can in this case be on the boundary of $S(\Gamma)$, for $S(\Gamma)$ is clearly contained in $\mathfrak{D}^*(C) \cup C \cup \mathfrak{D}^\#(C)$, a complementary domain of $\delta(C)$, whereas every other curve of T_C lies in one or two other complementary domains of $\delta(C)$. (Note: $\delta(C)$ divides π into at most three Jordan domains.

On the other hand, suppose that C is a curve of F on the boundary of $S(\Gamma)$. (Note: from what follows it is clear that the boundary is indeed a union of curves of F.) Then, directing C so that $\mathfrak{D}^*(C)$ contains the initial point of Γ, we note that if p is a point on C, limit point of a sequence p_n of $S(\Gamma)$, then there is an r-set $U(pq)$ of any arc pq on C^* and a sequence $q_n \rightarrow q$ with $q_n \varepsilon C_{p_n}$, $C_{p_n} \subset S(\Gamma)$, from which we conclude that q is either in $S(\Gamma)$ or on its boundary. If C^* does not cross Γ, then q will be on the boundary and C^* is a boundary curve of $S(\Gamma)$. When this is the case, C^* divides π into two domains $\mathfrak{D}^*(C^*) \supset S(\Gamma)$ and $\mathfrak{D}^\#(C^*) \supset T_C - C^*$, whence no other points of T_C other than those of C^* are on the boundary of $S(\Gamma)$. But, if C^* crosses Γ at a point p on a curve C', then we are back in the previous case and $\delta(C') = [C^* \cup C^\#] - C'$ is the boundary in T_C of $S(\Gamma)$.

THEOREM 2. 2-2. *Let q be a point on a curve C_q of $F^* = F[R^*]$ and let $\Gamma(q)$ be a cross-section tending properly to infinity in R^* in each direction. Further, let h be any homeomorphism of R^* onto the xy-plane, then $h[\Gamma(q)]$ is a cross-section of the family $h[F^*]$ (filling the xy-plane) which tends properly to infinity in both directions on the xy-plane.*

Proof. On the xy-plane we let K_n denote a circle of radius n, center $h(q)$. We must show that for every n there are points q'_n, r'_n on $\Gamma' = h[\Gamma(q)]$ such that every curve of $h(F^*)$ intersecting Γ' at points outside the arc $q'_n r'_n$ will lie outside K_n. If this is not the case, as we shall assume, we will be able to find a sequence of points t'_n receding to infinity on Γ' such that each $C_{t'_n}$ intersects a fixed circle K_N of the circles K_n. Now the inverse image K of K_N is a simple closed curve in R^* containing q in its interior. We will denote by C_n the inverse image of $C_{t'_n}$ and by t_n the inverse image of t'_n. Every C_n must then intersect K and hence intersect some circle with center at q which contains K. But this contradicts the assumption that $\Gamma(q)$ tended properly to infinity in R^*, since we have a sequence t_n approaching infinity on $\Gamma(q)$, but the curves C_{t_n} do not approach infinity. Hence the theorem must be true.

W. Kaplan introduced the notion of admissible collections of finite sequences in order to number the half-parallel subsets of a regular curve family filling an open simply connected domain. The concept is so similar to that already considered in the numbering of curves of a tree that we shall be able to use the same notation as in that section. As in Kaplan [II], we shall call a collection A of finite sequences *admissible* if

(1) A contains the one-element sequence 1 and no other one-element sequences, and

(2) $\alpha, k \varepsilon A$ implies $\alpha, k-1 \varepsilon A$ if $k > 1$ and implies $\alpha \varepsilon A$ if $k = 1$.

Now, if we have a regular curve family F' filling the xy-plane, and if we have assigned to each point (x,y) an extended cross-section $\Gamma(x,y)$ tending properly to infinity in both directions, then for any fixed curve C_1 it was shown in [II] that we can decompose $F'[C_1 \cup \mathfrak{D}^*(C_1)]$ into a collection of non-overlapping, half-parallel subfamilies $S(\alpha)$ which will be numbered by the finite sequences $\{\alpha\}$ of an allowable collection A. Each half-parallel family $S(\alpha)$ will be the set of all curves intersecting a cross-section $\Gamma(\alpha)$ tending from an initial curve C_a to infinity and lying on some $\Gamma(x,y)$ as chosen above; C_a will be the only curve of $S(\alpha)$ mapped onto the x-axis in the homeomorphism of $S(\alpha)$ onto the lines $y = k \geqq 0$ and the complete boundary of $S(\alpha)$ will be, in addition to C_a, just exactly the curves $C_{a,k}$. Note that when we write C_a we mean to indicate that C_a is an initial curve of some $S(\alpha)$ in the decomposition of F', whereas $C(\alpha)$ will, as in [I], indicate that C is the curve of a numbered tree which has been assigned the signed sequence α in the numbering of the tree.

As a corollary to the preceding Theorem 2.2-2 plus the proof of the facts mentioned in the preceding paragraph from [II], we can immediately state the following theorem:

THEOREM 2.2-3. *Given the family $F^* = F[R^*]$ and an arbitrary regular curve C_1 of F^*, we can decompose $F^*[C_1 \cup \mathfrak{D}^*(C_1)]$ (which is the same as $F[C_1 \cup \mathfrak{D}^*(C_1) \cap R^*]$) into a collection of non-overlapping half-parallel subsets $S(\alpha)$, each $S(\alpha)$ being all curves intersecting a cross-section $\Gamma(\alpha)$ tending from a curve C_a in F^* properly to infinity in R^*.*

In order to study the relation between an arbitrary tree T of F and a given decomposition of F^* into sets $S(\alpha)$ ($\alpha \varepsilon A$, as described above), it is convenient to adopt some new notation. $A(T)$ will denote the subset of A containing all sequences α such that $S(\alpha) \cap T \neq 0$; and $A_n(T)$ the subset of all sequences of $A(T)$ of order n. We denote by $N(T)$ the smallest integer n such that $A_n(T)$ is not empty. It is clear that $\Gamma(\alpha)$ can have at most one point on T, and hence $S(\alpha) \cap T$ is a curve of F^* or is empty. If $\Gamma(\alpha) \cap T$ is the initial point of $\Gamma(\alpha)$ we say that $\Gamma(\alpha)$, or $S(\alpha)$, *begins at* T; in this case $C_a = S(\alpha) \cap T$. When $\Gamma(\alpha) \cap T$ is a point of $\Gamma(\alpha)$ other than the initial point, then $\Gamma(\alpha)$, or $S(\alpha)$ is said to *straddle* T. In the former case $S(\alpha)$ lies in one complementary domain of T, in the latter in two. Using these notations, we may state the following properties:

(1) If α, β are *distinct* elements of A with $\beta \varepsilon A(T)$, and α either an element of $A(T)$ or such that points of T lie on the boundary of $S(\alpha)$; then $S(\alpha)$, $S(\beta)$ cannot each intersect the same complementary domain of T.

(2a) If $A_N(T)$, $N = N(T)$, has one element α, then either $S(\alpha)$ straddles T, or if $S(\alpha)$ begins at T, then $C_{\alpha}{}^{*} \cap R^{*} = S(\alpha) \cap T$, i. e., $C_{\alpha}{}^{*}$ has just *one* curve in R^{*}.

(2b) If $A_N(T)$ has more than one element, then every element of $A_N(T)$ is of the form β, k for fixed β of order $N - 1$ and $C_{\beta,k}$ for $\beta, k \varepsilon A_N(T)$ are just those curves of a maximal chain C^{*} which are in R^{*}.

(3) Let γ be an element of $A_{N+k}(T)$, then every lower segment of γ of order $\geqq N(T)$ is in $A(T)$, i. e., for $0 \leqq j \leqq k$ we have $\gamma_{N+j} \varepsilon A_{N+j}(T)$, where, as previously, γ_{N+j} is the sequence consisting of the first $N + j$ elements of the sequence γ.

(4) A necessary condition that $S(\alpha)$ straddle T is that $\alpha \varepsilon A_N(T)$ and is the only element of $A_N(T)$.

First we prove (1). Let $\mathfrak{D}^{*}(C)$ be a complementary domain of T, bounded by C^{*} on T. Suppose that $S(\alpha)$ and $S(\beta)$ both have points in $\mathfrak{D}^{*}(C)$. Then there is a point p_1 on $\Gamma(\alpha)$, p_2 on $\Gamma(\beta)$, each in $\mathfrak{D}^{*}(C)$. Now since $\beta \varepsilon A(T)$, $\Gamma(\beta)$ has a point q_2 on C^{*} and p_2q_2, an arc on $\Gamma(\beta)$, lies in $\mathfrak{D}^{*}(C) \cup C^{*}$. In either of the possibilities for α mentioned above, there would be a point q_1 on C^{*} which was a limit point of points q'_n in $S(\alpha)$. If $\alpha \varepsilon A(T)$ then q'_1 and q'_n may be taken on $\Gamma(\alpha)$, otherwise q'_n will consist of points in $\mathfrak{D}^{*}(C)$ which are in $S(\alpha)$ but not on $\Gamma(\alpha)$. It follows by arguments used many times above, i. e. considering an r-set of q_1q_2, etc., that there is a cross-section from q_1 on C^{*} into $\mathfrak{D}^{*}(C)$, which may be shown to cross a curve also crossed by p_2q_2. This curve would have to be in both $S(\alpha)$ and $S(\beta)$ which is impossible since α, β were assumed distinct. The following lemma will be useful in proving properties 2-4.

Lemma. *If $\alpha \varepsilon A(T)$ and $\alpha, k \not\varepsilon A(T)$, then no sequence γ of $A(T)$ can have α, k as a lower segment.*

Proof. $C_{\alpha,k}$ lies on the boundary of $S(\alpha)$ but is not in T, nor is any curve of $S(\alpha, k)$ in T by hypothesis. Assuming $C_{\alpha,k}$ directed so that $S(\alpha, k) \subset \mathfrak{D}^{*}(C_{\alpha,k})$, we have by Theorem 2.2-1 two possibilities: (a) the entire curve $C^{\#}_{\alpha,k}$ is on the boundary of $S(\alpha)$ and is all of this boundary in the tree $T' = T_{C_{\alpha,k}}$, or (b) T' intersects $S(\alpha)$ on a curve C of T', and then $C_{\alpha,k} \subset \delta(C)$ where $\delta(C)$ is on the boundary of $S(\alpha)$ and is all of this boundary in T'. In case (a) every curve of $C^{\#}_{\alpha,k} \cap R^{*}$, being on the boundary of $S(\alpha)$, is a curve $C_{\alpha,k'}$ for some k'. We have $\mathfrak{D}^{\#}(C^{\#}_{\alpha,k}) \supset T$, since it contains $S(\alpha)$ which intersects T. In case (b) $\delta(C) = \delta(C+) \cup \delta(C-)$ divides π into three domains (or two if one of the sets $\delta(C \pm)$ is empty);

one of these which we denote D_1 contains C and hence $S(\alpha)$ and T. The others, D_2 and D_3, contain all other curves of T'. $\delta(C)$ is the complete boundary in T' of $S(\alpha)$, hence every curve of $\delta(C) \cap R^*$ is a curve $C_{\alpha,k'}$ for some k'.

The remainder of the proof depends on the fact that $S(\beta) \cup S(\beta, k)$ is always a connected set. If there exists any sequence $\gamma = \alpha, k, n_1, \cdots, n_r$ such that γ is in $A(T)$ then, $S(\gamma)$ must clearly have points in $\mathfrak{D}^{\#}(C^{\#}_{\alpha,k})$ above in case (a) or in D_1 in case (b), these being the domains of T' in which T lies, and hence, so also has the set $\Sigma = \bigcup_{j=0}^{r} S(\alpha, k, n_1, n_2, \cdots, n_j)$. Moreover, the set Σ is connected, and $S(\alpha, k)$ which is in this set lies in $\mathfrak{D}^{*}(C_{\alpha,k})$ in case (a), and in D_2 or D_3 in case (b). Thus Σ has points on either $C^{\#}_{\alpha,k}$ or $\delta(C)$, the boundary curve, i. e., for a $j \neq 0$ there is a curve of $C^{\#}_{\alpha,k}$ or $\delta(C)$ as the case may be in $S(\alpha, k, n_1, \cdots, n_j)$. But each such curve as already pointed out is a curve $C_{\alpha,k'}$, which is a contradiction.

The lemma implies in particular, that if α and $\alpha, n_1, \cdots, n_r \,\varepsilon\, A(T)$ then $\alpha, n_1, \cdots, n_j \,\varepsilon\, A(T)$, $j \leqq r$. Hence (3) will follow if we prove that every sequence of $A(T)$ contains a lower segment in $A_N(T)$.

Now we turn to an examination of the possibilities for $A_N(T)$ and completion of the proof of (3). Suppose that α is an element of $A_N(T)$. Then either $S(\alpha)$: (i) straddles T, or (ii) begins at T. In the former case let $C = S(\alpha) \cap T$, then $\delta(C)$ is the complete boundary of $S(\alpha)$ in T, and we know that every curve in R^* of $\delta(C)$ is in the collection $\{C_{\alpha,k}\}$. Moreover, $\delta(C)$ divides π into three (or two) domains $D_1, D_2, (D_3)$ of which the first contains C_1, and of T, *only* the curve C. Now let γ be any sequence of $A(T)$. $S(\gamma)$ must, by (1), lie in D_2 or D_3. But $\bigcup_{i=1}^{n} S(\gamma_i)$ is a connected set containing C_1 (i. e., C_α, $\alpha = 1$), hence points of D_1 and also points of D_2 or D_3. It must then contain a curve $C_{\alpha,k}$ of $\delta(C)$, and therefore $S(\alpha)$, i. e., α is a lower segment of γ. Since this is only possible if γ is of order $> N$ we conclude α is the only element of $A_N(T)$.

In the case (ii) where $S(\alpha)$ begins at T, we have C_α on the boundary of $S(\beta)$, where β is of order $N - 1$ and $\alpha = \beta, k$. In fact, $C^{\#}_\alpha$ is the complete boundary on T of $S(\beta)$ (the curves of $C^{\#}_\alpha \cap R^*$ are all in the set $\{C_{\beta,k'}\}$ for some k' and therefore are in $A_N(T)$); and we have $\mathfrak{D}^{\#}(C^{\#}_\alpha) \supset S(\beta)$, $\mathfrak{D}^{*}(C^{*}_\alpha) \supset S(\alpha)$. Now let us suppose that $\gamma \,\varepsilon\, A(T)$, then $S(\gamma)$ by (1) cannot lie in $\mathfrak{D}^{\#}(C^{\#}_\alpha)$, hence must lie in $\mathfrak{D}^{*}(C^{\#}_\alpha)$. But $\bigcup_{i=1}^{n} S(\gamma_i)$ is connected and has a point in $\mathfrak{D}^{\#}(C^{\#}_\alpha)$, namely, any point of C_1. Thus this set has a point on $C^{\#}_{\alpha} \cap R^{*}$ and hence contains a curve $C_{\alpha,k'}$. It follows that every sequence of $A(T)$ has a lower segment in $A_{N}(T)$. This proves (1) and completes the proof of (3).

To prove (4) we need show only that if $S(\alpha)$ straddles T, then no lower segment of α is in $A(T)$. If $\alpha = \beta, k$, so that β is the lower segment α_{n-1}, then if any lower segment of α is in $A(T)$, β is also by our lemma. Then C_{α}, being on the boundary of $S(\beta)$, we necessarily have $S(\beta)$, $S(\alpha)$ in different domains of $T_{C_{\alpha}}$. This is impossible unless $T = T_{C_{\alpha}}$, for we would otherwise have points of T in two different domains of $T_{C_{\alpha}}$. But then $S(\alpha)$ does not straddle T, for on the contrary $C_{\alpha} \subset T$.

As above we consider the branched regular curve family F with a regular curve C_1 of F and a decomposition of the corresponding $F^{*}[C_1 \cup \mathscr{D}^{*}(C_1)]$ into sets $S(\alpha)$ with initial curves $C(\alpha)$. Then using properties 1-4, we have the following:

THEOREM 2. 2-4. *The complete half-parallel subfamilies* $S^{*}(\alpha) = S(\alpha) \cup C^{*}_{\alpha}$ *decompose* $F[\mathscr{D}^{*}(C_1) \cup C_1]$ *into a family of half-parallel subsets which intersect only on a curve of their initial curves, i. e.,* $S^{*}(\alpha) \cap S^{*}(\beta) = 0$ *or* C *where* $C^{*} = C^{*}_{\alpha}$ *and* $C^{\#} = C^{\#}_{\beta}$. (See Figure 4.)

Proof. First to prove that every curve of $F[C_1 \cup \mathscr{D}^{*}(C_1)]$ is included in this decomposition we note that every curve of $F^{*}[C_1 \cup \mathscr{D}^{*}(C_1)]$ is automatically included, being already in a set $S(\alpha)$ of the decomposition of that part of the simply connected region R^{*} included in $\mathscr{D}^{*}(C_1)$. Thus we have only to consider curves of $\tilde{\mathscr{T}}$; let C be a curve of $F[C_1 \cup \mathscr{D}^{*}(C_1)]$ which is not in R^{*} and let T denote the tree which contains it. Then no cross-section $\Gamma(\alpha)$ has a point on C. C will be on the boundary of two distinct sets $S(\alpha)$ and $S(\beta)$ in $\mathscr{D}^{*}(C)$ and $\mathscr{D}^{\#}(C)$ respectively. They cannot coincide since if they did, then it would mean that $S(\alpha) = S(\beta)$ would straddle T, for otherwise the set $S(\alpha)$ lies in a single domain of T. But then, in this case, since the cross-section $\Gamma(\alpha)$ would have points in two domains both having C and only C as common boundary, it would have to contain a point of C, which is clearly impossible if C is not in R^{*}.

Now, if either α or β, say α, is of order $> N(T)$ then, since C, a curve of T is on the boundary of $S(\alpha)$, α, k for some k is in $A(T)$. Hence by (3), α must also be in $A(T)$. Then by (4), C_{α} must lie on T, whence we have at once that $C^{*}_{\alpha} = C^{*} \supset C$ and hence C is in $S^{*}(\alpha)$. Thus it remains to show that either α or β must be of order $> N$. Clearly each is of order $\geqq N - 1$ since, for example, curves of T are on the boundary of $S(\alpha)$ hence either $\alpha \varepsilon A(T)$ or $\alpha, k \varepsilon A(T)$ for some k. Assume α is of order $N - 1$, then $\Gamma(\alpha)$ does not have a point on T and by Theorem 2. 2-1 all of C^{*} is on the boundary of $S(\alpha)$, and every curve of $C^* \cap R^*$ is in the set $\{C_{\alpha,k}\}$, and for each of these $\alpha, k \,\varepsilon\, A_N(T)$. Now, since β, k' for some k' is in $A(T)$, β is of order $\geqq N-1$. If now β were of order $N-1$, then there must be a $\beta, k' \,\varepsilon\, A_N(T)$ and β, k' would for some k equal α, k by property (2b), whence $\beta = \alpha$ by (1); or if β were of order N, then it is of the form α, k for some k. This latter would mean that the common boundary of the domains containing $S(\alpha)$ and $S(\beta)$ would be the curve $C_{\alpha,k}$ which must then coincide with the curve C, contrary to assumption that C is not in R^*. Hence β in this case must be of order $> N$. On the other hand, if α is of order N, then either β is of order $> N$ or C_α and C_β lie on the same maximal curve $C^{\#}_\alpha = C^{\#}_\beta$, and by (2b) in this case quite clearly, C^*_α and C^*_β could not have a boundary curve C in common. Hence either α or β is of order $> N$ and we have already shown that in this case C is in the initial curve C^*_α or C^*_β of either $S^*(\alpha)$ or $S^*(\beta)$ respectively.

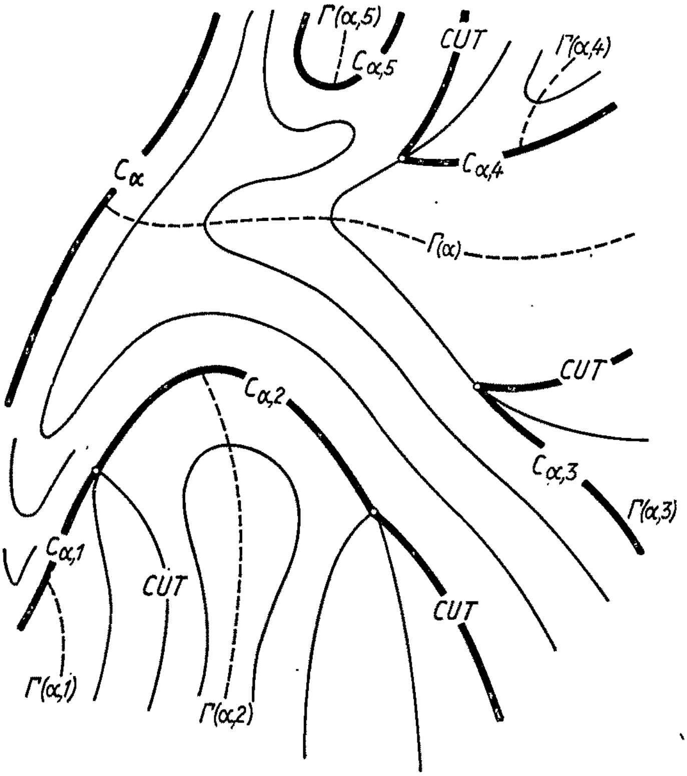


FIGURE 4

Next it must be shown that if C_α is the initial curve of a set $S(\alpha)$, then for any $S(\beta)$ which intersects C^*_α, the intersection must be C_β. Let C be the curve of intersection, i. e., $C = C^*_\alpha \cap S(\beta)$. Thus, $\alpha, \beta \varepsilon A(T)$ where T is the tree containing C_α. Now $S(\alpha)$ and $S(\beta)$ cannot have points in the same complementary domain of T by (1), which means in particular that $S(\beta)$ cannot straddle T, i. e., $\Gamma(\beta) \cap C =$ initial point of $\Gamma(\beta)$, since one complementary domain of C is $\mathfrak{D}^*(C^*_\alpha)$, C lying as it does on C^*_α, and $S(\alpha)$ lies in this complementary domain. Hence $C_\beta = C$ which was to be proved. From this it follows that two complete half-parallel subfamilies intersect only along a curve of their initial curves, and when they do intersect on a curve C, one lies in $\mathfrak{D}^*(C)$ and the other in $\mathfrak{D}^\#(C)$.

THEOREM 2. 2-5. *The family F can be decomposed into complete half-parallel subfamilies which overlap at most along their initial curves.*

Proof. We merely begin with any regular curve C_1 and decompose both $C_1 \cup \mathfrak{D}^*(C_1)$ and $C_1 \cup \mathfrak{D}^\#(C_1)$ as above.

We remark in conclusion that if $f(p)$ is a function with the family F as level curves, and if $S^*(\alpha)$ is a half-parallel subfamily of the decomposition, then clearly $f(p)$ is strictly monotone on $\Gamma(\alpha)$ and hence cannot assume the same value on two curves of $S^*(\alpha)$. If F is the level curve family of the real part of a single-valued analytic function, then clearly this function is 1-1 in $S(\alpha)$, i. e. as noted earlier the $\{S^*(\alpha), \alpha \varepsilon A\}$ give a decomposition of the domain of the analytic function into domains in which it is schlicht.

NORTHWESTERN UNIVERSITY.

BIBLIOGRAPHY.

[I] W. M. Boothby, "The topology of regular curve families with multiple saddle points," *American Journal of Mathematics*, vol. 73 (1951), pp. 405-438.

[II] W. Kaplan, "Regular curve-families filling the plane, I," *Duke Mathematical Journal*, vol. 7 (1940), pp. 154-185.

[III] ———, "Topology of level curves of harmonic functions," *Transactions of the American Mathematical Society*, vol. 63, No. 3 (1948), pp. 514-522.

[IV] ———, "Differentiability of regular curve-families on the sphere," *Lectures in Topology*, Ann Arbor (1941), pp. 299-301.

[V] S. Stoilow, *Leçons sur les principes topologiques de la théorie des fonctions analytiques*, Paris, 1938.

[VI] H. Whitney, "Regular families of curves," *Annals of Mathematics* (2), vol. 34 (1933), pp. 244-270.

[VII] G. T. Whyburn, *Analytic Topology*, American Mathematical Society Colloquium Publications, vol. 28, New York (1942).

FINITE METABELIAN GROUPS AND THE LINES OF A PROJECTIVE FOUR-SPACE.*

By H. R. BRAHANA.

---

1. Introduction. This paper presents a classification of the groups with the following properties:

(1) The groups are metabelian, i. e. of class 2;
(2) The elements are all of order p except the identity;
(3) In each group the central and coommutator subgroups coincide;
(4) Each group is generated by five elements.[1]

It was shown in the second paper just cited that there is a unique group of order p^{15} satisfying the above conditions, and that every other group satisfying them is a quotient group of this group of maximum order. Our method of classification is to determine possible types of invariant subgroup of the maximum group. The determination of types is accomplished by means of properties which are invariant under a change of generators. A geometric representation of subgroups of the maximum group enables us to make use of geometric invariants to do this.

Denote the maximum group by G and its generators by U_1, U_2, U_3, U_4, U_5. The commutator subgroup C is generated by the elements $c_{ij} = U_i^{-1}U_j^{-1}U_iU_j$, $i, j = 1, \cdots, 5$. c_{ii} is the identity; $c_{ji} = c_{ij}^{-1}$. Hence, C is of order p^{10}. Every element of G can be written cx where c is an element of C and $x = U_1^{x_1}U_2^{x_2}U_3^{x_3}U_4^{x_4}U_5^{x_5}$. The numbers x_i belong to GF(p). Two elements cx and $c'y$ have the same commutator as x and y, which is $\Pi c_{ij}^{p_{ij}}$, $i < j$, $p_{ij} = x_iy_j - x_jy_i$. If we represent the element x by the point $(x_1, x_2, x_3, x_4, x_5)$ in a finite projective four-space X, the commutator of x and y may be represented by the line xy and its expression in terms of commutators of pairs of generators of G will have for exponents the Plücker line-coordinates of xy. The Plücker coordinates of a line may be ordered and each line

---

* Received May 12, 1950.

[1] The relation of the problem of the groups having the first three properties to the general problem of groups of order p^m was discussed at considerable length in the author's paper "Metabelian groups and trilinear forms," *Duke Mathematical Journal*, vol. 1 (1935), pp. 185-197. Groups generated by four elements were classified in "Finite metabelian groups and Plücker line-coordinates," *American Journal of Mathematics*, vol. 62 (1940), pp. 365-379.

of the four-space may be represented by a point of a finite projective space S of nine dimensions. Every line in X determines a point in S but not every point in S corresponds to a line in X. Every point in S, however, determines a cyclic subgroup of C, necessarily of order p.

A set of defining relations of G contains the definitions of the c_{ij}'s given above and some other relations which were not given explicitly, viz.: $c' \cdot cx = cx \cdot c'$, which says G is metabelian; $cx \cdot c'y = c'y \cdot cx$ for all y only if $x_i = 0$, which says G is not the direct product of an abelian group and a metabelian group generated by fewer than five elements; and $\Pi c_{ij}^{\alpha_{ij}}$ is the identity only if $\alpha_{ij} = 0$ for all pairs i, j, which says that the c_{ij}'s are independent and G is the maximum group. Hence any group satisfying the conditions of the present problem will have defining relations the same as these except that the last will be replaced by one or more relations stating that one or more elements of the commutator subgroup is the identity. By a well-known theorem due to Dyck[2] this says that every group of the type we are considering is a quotient group of G with respect to some subgroup of C.

Every subgroup of C corresponds to some linear space in S, and every linear space in S determines a subgroup of C. Our problem then involves the determination of relations of all possible linear subspaces of S to the subset V of points of S which represent lines of X. A simple isomorphism of G with itself determines a collineation of the space X, a transformation of the space S into itself, and a transformation of V into itself. A simple isomorphism of G with itself which transforms U_i into $c^{(i)}U_i$, $i = 1, \cdots, 5$, determines the identity as the collineation in X and hence as the transformation in S; it determines the identity isomorphism also of the quotient group G/C. Isomorphisms of G with itself which are significant as far as S is concerned are those which determine non-identity isomorphisms of G/C with itself, and there is a one-to-one correspondence between these and the collineations of X. Two subgroups of C which are conjugate under the group of isomorphisms of G determine simply isomorphic quotient groups of G; their corresponding linear spaces in X are the same or are conjugate under a non-identity collineation of X. Hence we are interested in the conjugacy of linear spaces in S under collineations in X. These collineations may always be thought of as changes of generators of G/C.

In the following pages the types of point, line, plane, and three-space in S are given. For each type a canonical form is given. The canonical forms are all given for $p = 7$. One purpose of the canonical form is to establish the existence of the type; another purpose is to exhibit the relation to V of the corresponding subspace of S. These relations always come finally to the question of the reducibility or irreducibility of polynomials in one variable with coefficients in GF(p). The canonical form sometimes depends on the selection of a polynomial belonging to GF(7) having certain properties; the existence of a polynomial having the required properties nowhere depends on p being 7. For $p = 3$ or 5 some of the types may not exist or may coincide with others that are different from them for $p \geqq 7$; the argument is carried out in geometric terms and in a few places might run into difficulties if a conic has only four points or if a pencil has only four or six members.

---

[2] "Gruppentheoretische Studien," *Mathematische Annalen*, vol. 20 (1880), p. 14.

The grassmannian of lines in a four-space has been studied before, generally for the field of complex numbers. A paper[3] by J. A. Todd develops a wealth of geometry and gives many references to the literature. The interpretation of these results in a geometry based on GF(p) is in general obvious. In the third section of his paper Todd deals with linear spaces in S, with emphasis on their intersections with V. He determines points, lines, and planes; he gives some results about five-spaces which can be turned into theorems about three-spaces by means of the existing duality. The completeness of the list of planes which follows, and which was obtained independently, gives a verification of the completeness of Todd's list, which he does not state explicitly.

The completeness of the list of three-spaces which follows is not certain. Every S_3 in S determines a quintic polynomial in a single variable x. If the quintic is reducible, a coordinate system in X can be selected so that S_3 has one of the forms 1 to 53. These 53 types are distinguished from one another and from S_3's which give irreducible quintics by the presence in them of special points, lines, and planes. If the quintic is irreducible none of these special points, lines, or planes is present. Every irreducible quintic determines an S_3. The irreducible quintics belong to sets of conjugates under rational transformations of the variable x. An S_3 given by an irreducible quintic $f(x)$ may, by a change of coordinates in X, be made to give any quintic conjugate to $f(x)$, or it may be made to give a quintic which is not conjugate to $f(x)$; it is not certain that $f(x)$ can be changed to an arbitrary irreducible quintic $g(x)$. The type 54 can be characterized, and in many ways, by geometric properties which determine the coordinate system. An S_3

---

[3] "The locus representing the lines of four-dimensional space and its application to linear complexes in four dimensions," *Proceedings of the London Mathematical Society*, series 2, vol. 30 (1921), pp. 513-550.

which gives an irreducible quintic has determined in it by its relation to V a set of p^2+1 rational cubic curves no two of which intersect; the points of the cubics therefore fill S_3. The form 54 is determined by the presence in the set of two cubics properly related to each other. The pair is determined by either of the cubics. The relations that two cubics can have to each other fall into at most 21 sets, irrespective of p and irrespective of the irreducible quintic. It has been found that not every cubic can be a member of a "canonical" pair; it seems likely that every such set of cubics contains at least one canonical pair.

2. Some relations between elements in X and elements in S. In order to avoid double subscripts we adopt a notation as follows: let x and y be two points of the four-space X; then the line xy determines in S the point

$$(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}) = (p_{12}, p_{13}, p_{14}, p_{15}, p_{23}, p_{24}, p_{25}, p_{34}, p_{35}, p_{45}).$$

The manifold V is the set of points in S whose coordinates satisfy certain obvious relations among the two-rowed determinants p_{ij}. In terms of the a's they are $B_1=0, \cdots, B_5=0$, if $B_1, \cdots, B_5$ denote

$$a_1a_8 - a_2a_6 + a_3a_5, \qquad a_1a_9 - a_2a_7 + a_4a_5, \qquad a_1a_{10} - a_3a_7 + a_4a_6,$$
$$a_2a_{10} - a_3a_9 + a_4a_8, \qquad a_5a_{10} - a_6a_9 + a_7a_8,$$

respectively. Three of the five relations are independent, but all five are linearly independent.

Two lines of X determine two points of V. If the two lines of X intersect, it is well known[4] that the lines of the pencil defined by them are represented by the points of a line on V, and the lines in the plane determined by them are represented by the points of a plane on V.

There are obviously two kinds of points in S: (1) points on V; and (2) points not on V. Points on V are all alike since each represents a line X and there exists a collineation of X transforming any line into any other.

We proceed to prove that points of S not on V are also all of one type. First we prove the theorem:

Every point of S not on V is on a line joining two points of V.

Let $P=(a_1, \cdots, a_{10})$ be not on V. We may suppose that $a_{10} \neq 0$. Let us denote the vertices of the frame of reference in X by $A_1=(1,0,0,0,0)$, $A_2=(0,1,0,0,0), \cdots, A_5=(0,0,0,0,1)$. If we change the coordinate system in X to one that has the same vertices except that $A'_4 = (a_4, a_7, a_9, a_{10}, 0)$ and make the corresponding changes in coordinates in S, then P will become $(a'_1, a'_2, a'_3, 0, a'_5, a'_6, 0, a'_8, 0, 1)$. This is a point of the line joining $(a'_1, a'_2, 0, 0, a'_5, 0, 0, 0, 0, 0)$ and $(0, 0, a'_3, 0, 0, a'_6\ 0, a'_8, 0, 1)$, and both of these points are on V. These points represent lines in X which are skew, since P is not on V. We may select a new frame of reference in X with two vertices on each of these two lines. If we take A_1 and A_2 on one and A_3 and A_4 on the other, the point P becomes $(1, 0, 0, 0, 0, 0, 0, 1, 0, 0)$.[5]

---

[4] The two lines in X are in a three-space in X and we apply Theorem 30, Veblen and Young, *Projective Geometry*, vol. 1, p. 329.

We have shown the existence of a line on two points Q and Q' of V through any point P not on V. Q and Q' are images in S of two lines q and q' in X. The two skew lines q and q' determine a three-space in X. A second line on P which intersects V in two points determines a three-space in X. We proceed to prove that these two three-spaces coincide. We suppose P to have the form just obtained; we take a point $Q = (a_1, a_2, \cdots, a_{10})$ on V, and find the condition on the a's in order that the line PQ meet V in two distinct or coincident points. Expressing the coordinates of a point on PQ in parametric form in terms of the coordinates of P and Q and substituting in the equations of V we obtain the conditions for two solutions to be that $a_4 = a_7 = a_9 = a_{10} = 0$. This says that Q must be a point on V which corresponds to a line in the three-space $x_5 = 0$ of X. Hence,

Every point of S not on V determines a unique three-space R in X. The lines of this three-space are the only lines in X such that their corresponding points on V can be joined to the given point by a line that cuts V twice or is tangent to V.

The lines of a three-space R of X determine points of V which lie in a five-space Σ of S. Hence,

Every point of S not on V determines a unique five-space of S.

If $P = (a_1, a_2, \cdots, a_{10})$ is a point of S not on V, the numbers $B_1, B_2, \cdots, B_5$ defined above are not all zeros. We next prove the following theorem:

If P is a point of S not on V, the three-space R in X determined by P is $B_5x_1 - B_4x_2 + B_3x_3 - B_2x_4 + B_1x_5 = 0$.

Let $x = (x_1, \cdots, x_5)$ and $y = (y_1, \cdots, y_5)$ be two points of X that satisfy the above equation, and let the point of V determined by the line xy be $Q = (b_1, b_2, \cdots, b_{10})$. The condition that $Q + \lambda P$ be on V gives five quadratic equations in λ. Under the conditions on x and y these equations have solutions. Since Q is on V, $\lambda = 0$ is a solution of the five quadratics. Removing the factor λ the five quadratics become the five linear equations

---

[5] The question as to whether the two non-zero coordinates of P are the same or different is connected in the geometry with the selection of the unit point in X, and in the groups with the selection of a particular element in a cyclic group for a generator.

$$\sum_j b_j \partial B_i / \partial a_j - \lambda B_i = 0, \qquad i = 1, \cdots, 5.$$

The fact that x and y satisfy the equation above gives the relations:

$$B_4 b_1 - B_3 b_2 + B_2 b_3 - B_1 b_4 = 0, \qquad B_5 b_1 - B_3 b_5 + B_2 b_6 - B_1 b_7 = 0,$$
$$B_5 b_2 - B_4 b_5 + B_2 b_8 - B_1 b_9 = 0, \qquad B_5 b_3 - B_4 b_6 + B_3 b_8 - B_1 b_{10} = 0.$$

Making use of these it is easily verified that the λ given by one of the linear equations satisfies the others.

Consider two points P_1 and P_2 in S, neither on V. Each point determines a three-space R_i in X and a five-space Σ_i in S. If R_1 and R_2 are the same, then the line P_1P_2 lies in the five-space Σ_1 $(= \Sigma_2)$. Every point of the line P_1P_2 determines the same three-space R_1 $(= R_2)$ in X; a point of intersection of P_1P_2 with V is the image of a line in R_1. A line, plane, or three-space in S which lies in a five-space Σ determined by the lines of a three-space R in X will be called a *Σ-line*, a *Σ-plane*, or a *Σ-three-space*.

In dealing with Σ-spaces in S we may forget about the four-space X and retain only the three-space R. The five-space Σ intersects V in the four-dimensional hyperquadric determined by the lines of R. A Σ-line may be a ruling, a tangent, a secant, or it may have no point on the hyperquadric.[6] A ruling, a tangent, or a line with two points on V is always a Σ-line.

If the three-spaces R_1 and R_2 determined by P_1 and P_2 are distinct, then P_1P_2 is not a Σ-line. R_1 and R_2 will have a plane σ in common. The lines of σ determine points of a plane π on V, and π is in Σ_1 and Σ_2. Since π and P_1 are both in Σ_1, the polar of P_1 with respect to the intersection of Σ_1 and V intersects π in a line l_1. Likewise, the polar of P_2 with respect to the intersection of Σ_2 and V intersects π in a line l_2. The lines l_1 and l_2 may be distinct, or they may not.

If l_1 and l_2 are distinct, they intersect in a point M which is on V since it is in π. The lines joining M to P_1 and P_2 are both tangent to V at M. Consequently, the plane P_1P_2M is tangent to V at M, and every line in it through M is tangent to V. The point M is uniquely determined by P_1 and

---

[6] The condition for intersection of the line and V may lead to a quadratic irreducible in GF(p).

P_2; it is the only point of V from which tangents can be drawn to both P_1 and P_2. However, if P_1 and P_2 are moved along the line joining them, M remains unchanged. M is the image of a line m in X; m is in both R_1 and R_2. If P_1 is left fixed and P_2 is moved along the line P_1P_2, R_1 remains fixed but R_2 changes, and σ moves through the pencil of planes in R_1 on m. Thus the line m, and its image M on V, belongs to the line P_1P_2 and not to the pair of points P_1 and P_2.

We consider further the line P_1P_2. R_1 and R_2 are distinct and so are l_1 and l_2. Let Q_1 be a point other than M on l_2. The line P_1Q_1 is in Σ_1; it is not tangent to V at Q_1 since Q_1 is not on l_1. Hence, P_1Q_1 intersects V in a second point Q'_1. Likewise, if Q_2 is selected on l_1 different from M, P_2Q_2 is in Σ_2 and intersects V in a second point Q'_2. The plane $P_1Q_1Q_2$ is in Σ_1 since π is in Σ_1; hence it intersects V in a conic which is degenrate since it contains the line Q_1Q_2 and consists of two intersecting lines since Q'_1 is not on Q_1Q_2. The vertex of the conic is Q_2 since Q_2 is on l_1 and P_1Q_2 is a tangent. The line $Q_2Q'_1$ is therefore a line on V and hence the lines q_2 and q'_1 in X intersect. Similarly, the lines q_1 and q'_2 intersect. Also, the line m intersects both q'_1 and q'_2. The five lines q_1, q'_1, q_2, q'_2, and m have the following relations:

q_1, q_2, and m form a triangle in σ;
q'_1 intersects q_2 and m, is not in R_2, but is in R_1;
q'_2 intersects q_1 and m, is not in R_1, but is in R_2.

We select a frame of reference in X as follows:

A_1 is the intersection of q_1 and q_2;
A_2 is the intersection of q_1 and m, and is on q'_2;
A_3 is the intersection of q_2 and m, and is on q'_1;

A_4 and A_5 are on q'_1 and q'_2 respectively, and not in σ. With this coordinate system, we have in S

$$Q_1 = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0), \qquad Q_2 = (0, 1, 0, 0, 0, 0, 0, 0, 0, 0),$$
$$Q'_1 = (0, 0, 0, 0, 0, 0, 0, 1, 0, 0), \qquad Q'_2 = (0, 0, 0, 0, 0, 0, 1, 0, 0, 0),$$

$M = (0, 0, 0, 0, 1, 0, 0, 0, 0, 0)$. With a proper choice of the unit point, we have

$$P_1 = (1, 0, 0, 0, 0, 0, 0, 1, 0, 0), \qquad P_2 = (0, 1, 0, 0, 0, 0, 1, 0, 0, 0).$$

We may denote the line P_1P_2 by $k\,l\,0\,0\,0\,0\,l\,k\,0\,0$. We are thus introducing corrdinates k and l on the line, using P_1 and P_2 for vertices of the frame of reference on the line. An arbitrary point of the line determines the B's: $B_1 = k^2$, $B_2 = -l^2$, $B_3 = B_4 = 0$, $B_5 = kl$. A point of the line on V would require $k^2 = l^2 = 0$.

A second possibility above was that the points P_1 and P_2 were such that R_1 and R_2 were distinct, but such that the lines l_1 and l_2 *in* π coincided. In that case, if M is any point of l_1, then the plane MP_1P_2 is tangent to V at M. In order to investigate this we may take l_1, which is any ruling of V, to consist of the points

$$M_0 = (0, 1, 0, 0, 0, 0, 0, 0, 0, 0), \qquad M_\lambda = (\lambda, 1, 0, 0, 0, 0, 0, 0, 0, 0),$$

$M_\infty = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0)$. The spaces tangent to V are: at M_0, $a_6 = a_7 = a_{10} = 0$; at M_λ, $-a_6 + \lambda a_8 = 0$, $-a_2 + \lambda a_9 = 0$, $a_{10} = 0$; at M_∞, $a_8 = a_9 = a_{10} = 0$. If P_1P_2 is in two of these tangent spaces, then every point of it must have $a_6 = a_7 = a_8 = a_9 = a_{10} = 0$. Such a line is then P_1P_2, where

$$P_1 = (a_1, a_2, a_3, a_4, a_5, 0, 0, 0, 0, 0); \qquad P_2 = (b_1, b_2, b_3, b_4, b_5, 0, 0, 0, 0, 0).$$

The point $kP_1 + lP_2$ is on V for $a_5k + b_5l = 0$. Since P_1P_2 is by hypothesis not a Σ-line it can have no more than one point on V. We may then state the theorem:

The lines of S which are not Σ-lines belong to two types which are distinguished by the properties that a line of one type has no point on V, a line of the other type has one point on V. A line of the first type determines a unique point on V, a line of the second type determines a uniqwe line on V.

The space tangent to V at a point M of V is six-dimensional. It may be verified easily that a Σ-line is always in more than one tangent space, and thus every line of S is in the tangent space at some point of V. The space tangent to V contains planes, three-spaces, etc.; such will be called τ-*planes*, τ-*three-spaces*, etc. As will appear a little later, τ-planes and Σ-lines will be very useful in distinguishing among space in S. It is clear that the property of being a Σ-line, or a τ-plane, is invariant under collineations of X.

3. Types of points, of lines, and of planes in S.

The points of S.

1. 1 0 0 0 0 0 0 0 0 0, a point of V.
2. 1 0 0 0 0 0 0 1 0 0, a point not on V.

The lines of S.

1. $k\ l\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0$, a ruling of V.
2. $k\ 0\ 0\ 0\ 0\ 0\ 0\ l\ 0\ 0$, a secant of V.
3. $k\ 0\ 0\ 0\ l\ 0\ 0\ k\ 0\ 0$, a tangent to V.
4. $k\ l\ 0\ 0\ 0\ rl\ 0\ k\ 0\ 0$ (r not a square), a Σ-line with no point on V.
5. $k\ 0\ 0\ l\ 0\ 0\ 0\ k\ 0\ 0$, not a Σ-line, one point on V.
6. $k\ l\ 0\ 0\ 0\ 0\ l\ k\ 0\ 0$, not a Σ-line, no point on V.

The planes[7] *of S.*

We list first the six types of Σ-plane; all are in the five-space Σ determined by the three-space $x_5 = 0$ in X.

1. $k\ l\ 0\ 0\ m\ 0\ 0\ 0\ 0\ 0$, the image of a plane of lines in X. (xvi)
2. $k\ l\ m\ 0\ 0\ 0\ 0\ 0\ 0\ 0$, the image of a bundle of lines in X. (xv)
3. $k\ l\ 0\ 0\ 0\ 0\ 0\ m\ 0\ 0$, intersecting V in two lines. (xiii)
4. $k\ l\ m\ 0\ 0\ 0\ 0\ k\ 0\ 0$, intersecting V only in the line $k = 0$. (xiv)
5. $k\ l\ 0\ 0\ 0\ m\ 0\ k\ 0\ 0$, intersecting V in the conic $k^2 - lm = 0$. (xii)
6. $k\ l\ m\ 0\ 0\ rl\ 0\ k\ 0\ 0$, ($r$ not a square), intersecting V only in the point $P_3 : k = l = 0$. (xiii)

The rest of the planes belong to no Σ. We group them in sets according to their intersections with V.

Planes with no point on V.

7. $k\ l\ 0\ 0\ 0\ m\ l\ k{+}rm\ m\ 0$ ($x^3 + rx - 1$ irreducible), a τ-plane. (v)
8. $k\ l\ m\ 0\ {-rm}\ 0\ l\ k\ 0\ 0$ (r not a square), not a τ-plane but contains the Σ-line $l = 0$. (iii)
9. $k\ l\ 0\ 0\ m\ 0\ l\ k\ 0\ m$, not a τ-plane and contains no Σ-line. (i)

Planes with at least one point on V, but with no line on V.

10. $k\ l\ 0\ 0\ m\ 0\ l\ k\ 0\ 0$, tangent to V at P_3, one point on V. (viii)
11. $k\ l\ 0\ 0\ 0\ m\ l\ k\ 0\ 0$, a τ-plane, one point on V, one line tangent to V. (vii)
12. $k\ l\ m\ 0\ 0\ 0\ l\ k\ 0\ 0$, not a τ-plane, one point on V, one line tangent to V. (iv)
13. $k\ l\ 0\ 0\ 0\ m\ l{-}rm\ k\ 0\ 0$ (r not a square), a τ-plane, one point on V, no line tangent to V. (vi)

---

[7] The roman numbers locate these planes in Todd's list.

14. $k\ l\ m\ rm\ 0\ 0\ l\ k\ 0\ 0$, not a τ-plane, one point on V, no line tangent to V. (ii)
15. $k\ l\ 0\ 0\ 0\ 0\ l\ k\ 0\ m$, not a τ-plane, one point on V, no line tangent to V. (ii)
16. $k+m\ l\ 0\ 0\ 0\ 0\ l\ k\ 0\ 0$, a τ-plane, two points on V. (v)
17. $k+m\ l\ m\ 0\ 0\ 0\ l\ k\ 0\ 0$, not a τ-plane, two points on V. (iii)
18. $k+m\ l+m\ 0\ 0\ 0\ 0\ l\ k\ 0\ 0$, three points on V. (v)

Planes with a line on V.

19. $k\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ l\ m$, a line and an additional point on V. (x)
20. $k\ 0\ 0\ 0\ l\ 0\ 0\ k\ m\ 0$, a line on V and tangent at a point of it. (xi)
21. $k\ 0\ 0\ 0\ 0\ 0\ l\ k\ m\ 0$, a τ-plane, one line on V, not tangent at any point of the line. (ix)
22. $k\ 0\ 0\ 0\ 0\ 0\ 0\ k\ l\ m$, not a τ-plane, one line on V. (ix)

Each of the spaces in the above list is given in terms of a particular coordinate system, but with each is given a set of properties sufficient to differentiate among the spaces and these properties are not dependent on any coordinate system. Planes 14 and 15 provide an exception to the above statement, and we shall determine some geometric properties that will enable us to distinguish between them.

For plane 14, B_1, B_2, B_3, B_4, B_5 become k^2, $-l^2$, $-lm$, rkm, kl, respectively, while for plane 15 they become k^2, $-l^2$, km, lm, kl. In each case the conditions for a point on V are $k=l=0$, and hence each plane intersects V only at the point P_3. Each plane contains the line $P_1P_2 = k\ l\ 0\ 0\ 0\ 0\ l\ k\ 0\ 0$. P_1P_2 is in the space tangent to V at the point $M = 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0$; this tangent space is $a_3 = a_4 = a_{10} = 0$ which contains no other point of either plane. Hence, neither plane is a τ-plane. In plane 14 the point P_3 is $0\ 0\ 1\ r\ 0\ 0\ 0\ 0\ 0\ 0$ and the space tangent to V at P_3 is: $a_5 = 0$, $ra_6 - a_7 = 0$, $ra_8 - a_9 = 0$. The intersection of this tangent space with the plane requires $k = l = 0$, and hence it intersects the plane only at P_3. The space tangent to V at P_3 in plane 15 is $a_1 = a_2 = a_5 = 0$, and it intersects plane 15 only at P_3. So that the only difference between planes 14 and 15 might conceivably be due entirely to the choice of coordinate systems in the two cases. To show that this is not the case, we examine some new geometric relations between S and X.

We recall that if P is a point of S not on V, the numbers $B_1, \cdots, B_5$ are not all zero and that $B_5x_1 - B_4x_2 + B_3x_3 - B_2x_4 + B_1x_5 = 0$ is the three-space R in X determined by P. We now think of the above equation in another way. Let us suppose that $A = (x_1, x_2, x_3, x_4, x_5)$ is a given point in X. Then the equation becomes a quadratic in the coordinates $a_1, \cdots, a_{10}$ of a point in S. This eight-dimensional hyperquadric in S intersects a plane in S in a conic or else contains the plane completely. Thus every point of X determines in a plane π of S a conic, which may be identically zero. In the cases of planes 14 and 15 the conics given by A are

$$x_1kl - x_2rkm - x_3lm + x_4l^2 + x_5k^2 = 0,$$
$$x_1kl - x_2lm + x_3km + x_4l^2 + x_5k^2 = 0,$$

respectively. There is thus determined in each of the planes a four-parameter set W of conics. The points of a line in X determine a pencil of conics belonging to W. Each of planes 14 and 15 determines a special line in X, namely, the line p_3 which is imaged in S on the point P_3 common to the plane and V. Thus in each case the set W of conics contains a special pencil. For 14 the line p_3 joins A_1 and $A_4 + rA_5$ and the special pencil of conics is $\lambda kl + l^2 + rk^2 = 0$. For 15 the line p_3 joins A_4 and A_5 and the special pencil of conics is $\lambda l^2 + k^2 = 0$. Since $\lambda^2 - 4r = 0$ has no solution in GF(p), every conic in the special pencil in 14 consists of a pair of distinct lines. In the case of 15 $\lambda = 0$ and $\lambda = \infty$ give the conics $k^2 = 0$ and $l^2 = 0$. This proves that no change of coordinates will put 14 into 15.

It is comparatively easy to see that the planes listed above have the properties that are listed with them. This comes from the fact that use has been made of special elements in determining the canonical forms. Plane 9 is identified by its lack of special elements of the types that are used for the others. This plane does, however, contain a special element and the given form depends on proper use being made of it.

In the case of plane 9, the five conics given by the vertices of the frame of reference in X are linearly independent. Five linearly independent conics in a plane determine a unique conic which is apolar to all the conics of the four-parameter set W dependent on them. In plane 9 this conic is $C: m^2 - 2kl = 0$. P_1 and P_2 are points of C and P_3 is the pole of the line P_1P_2 with respect to C. Conversely, if π is any plane of S which has no point on V, which contains no Σ-line, and which is not a π-plane, and if P_1, P_2, and P_3 are selected to have the above relations to C, then a coordinate system in X can be selected so that π has the form 9.

4. Three-spaces in S. We separate the three-spaces of S into sets according to their intersections with V. There is one type of S_3 in S which lies wholly on V; it consists of the points which are images in S of all the lines in X through a single point. This S_3 is of no use in connection with the groups since it corresponds to the direct product of a cyclic group of order p and a metabelian group generated by four elements. We omit this S_3 from the list.

Three-spaces with at least a non-degenerate conic on V.

1. $k\,0\,m\,0\,n\,0\,0\,l\,0\,0$, the ruled quadric $kl + mn = 0$ on V.
2. $k\,l\,m\,0\,n\,{-l}\,0\,k\,0\,0$, the quadric $k^2 + l^2 + mn = 0$ on V; the quadric has no rulings.
3. $k\,l\,n\,0\,0\,l\,0\,m\,0\,0$, the cone $km - l^2 = 0$ on V.
4. $n\,n\,0\,0\,k\,m\,0\,0\,m\,l$, the conic $kl - m^2 = 0$, $n = 0$, and the line $l = m = 0$ on V.
5. $k\,l{+}n\,m\,0\,n\,n\,l\,k{+}n\,0\,0$, the conic $k^2 + kn - n^2 + mn = 0$, $l = 0$ and the point $(1\,1\,0\,{-1})$ on V.
6. $n\,0\,0\,0\,k\,m\,0\,n\,m\,l$, the conic $kl - m^2 = 0$, $n = 0$, on V.

Three-spaces with a plane on V, not lying wholly on V.

7. $k\,l\,m\,0\,n\,0\,0\,0\,0\,0$.
8. $k\,l\,m\,0\,0\,0\,n\,0\,0\,0$.
9. $k\,l\,m\,0\,n\,0\,0\,0\,0\,n$.
10. $k\,l\,0\,0\,m\,0\,0\,0\,0\,n$.
11. $k\,l\,0\,0\,m\,n\,0\,0\,n\,0$.

The planes on V are of two types: (1) planes whose points represent the lines of a plane in X, and (2) planes whose points represent the lines in a bundle. In spaces 7, 8, 9 above the plane $n = 0$ is of the second type; the first has two planes on V, the second has a plane and a line, and the third has only $n = 0$ on V. Spaces 10 and 11 have no plane of the second type; 11 is in the space tangent to V at P_3, while space 10 is not in the tangent space at any point of V.

Three-spaces with at least two rulings but no plane on V.

12. $k\,l\,0\,0\,0\,m\,0\,0\,0\,n$.
13. $k\,l\,0\,n\,0\,m\,n\,0\,0\,0$.
14. $k\,l\,0\,0\,0\,m\,0\,0\,n\,n$.
15. $k\,l\,0\,n\,0\,m\,0\,n\,0\,0$.
16. $k\,l\,0\,n\,0\,m\,n\,n\,0\,0$.
17. $k\,l\,0\,0\,n\,m\,0\,0\,0\,n$.
18. $k\,l\,0\,n\,n\,m\,0\,0\,0\,0$.

The first two have three rulings of V; in 13 the rulings pass through a point but in 12 they do not. Each of the others contains only two rulings and they intersect; 14 has an additional point on V and the others do not. 15 and 18 are in spaces tangent to V; 16 and 17 are not τ-spaces. 18 is in the space tangent to V at a point of its intersection with V; 15 is not.

To see that 16 and 17 are distinct types, we note that each contains two intersecting rulings of V and coordinate systems have been selected so that in both cases the line A_1A_2, in X, is imaged on this intersection. The line A_1A_2 is therefore special in both cases. An arbitrary point P in 16 determines the three-space R in X:

$$n^2x_1 - n^2x_2 + mnx_3 + lnx_4 + (kn - lm)x_5 = 0.$$

An arbitrary point in 17 gives the three-space

$$n^2x_1 - lnx_2 + knx_3 - lmx_5 = 0.$$

In the case of 16 every point P determines an R which contains the point 1 1 0 0 0, a point of the special line; in the case of 17 no point of the special line is in every R.

Three-spaces containing one ruling of V.

19. $m\ 0\ 0\ k\ 0\ 0\ l\ n\ k\ l$.
20. $k\ l\ 0\ 0\ 0\ n\ 0\ 0\ n\ m$.
21. $k\ l\ m\ n\ 0\ rn\ m\ 0\ 0\ 0$.
22. $k\ l\ 0\ 0\ 0\ n\ m\ m\ rn\ 0$.
23. $k\ l\ m\ 0\ m\ 0\ n\ n\ 0\ 0$.
24. $k\ l\ 0\ 0\ n\ 0\ m\ m\ 0\ {-rn}$.
25. $k\ l\ m\ 0\ n\ 0\ 0\ 0\ m\ n$.
26. $k\ l\ n\ 0\ 0\ 0\ m\ m\ n\ 0$.
27. $k\ l\ m\ n\ 0\ 0\ m\ n\ 0\ 0$.

In 21., 22., and 24., r is not a square.

Spaces 19 and 20 have respectively two and one points on V in addition to the ruling $m = n = 0$; the rest have only the ruling. Space 21 is in the space tangent to V at P_1; 22 is in the space tangent to V at a point not in 22; none of the others is a τ-space. In 23 every plane on P_1 is a τ-plane and every line in the plane $n = 0$ is tangent to V at its intersection with P_1P_2. Space 24 contains the τ-plane $n = 0$ and it contains a pencil of τ-planes on P_3P_4. P_3P_4 is a Σ-line which does not intersect the ruling. All τ-planes pass through P_3 but not every plane on P_3 is a τ-plane. Space 25 contains a pencil of τ-planes and no others; the axis of the pencil is P_2P_3, which is not the ruling. Space 26 contains a pencil of τ-planes on the ruling P_1P_2; one of the τ-planes is tangent to V at P_2. Space 27 contains a pencil of τ-planes on P_1P_2; one plane is tangent at P_1, another is tangent at P_2.

Three-spaces interesecting V in three or more points, but not in any plane curve.

28. $k{+}n\ k\ 0\ 0\ 0\ n\ l\ m\ n\ 0$.
29. $k\ k\ n\ {-n}\ {-n}\ 0\ l\ m\ 0\ n$.
30. $k\ k\ n\ {-n}\ 0\ 0\ l\ m\ 0\ n$.
31. $k\ k\ n\ n\ 0\ 0\ l\ m\ n\ n\ n$.
32. $k\ k\ n\ {-n}\ n\ 0\ l\ m\ 0\ n$.
33. $k\ k\ 0\ n\ n\ 0\ l\ m\ 0\ n$.
34. $k\ k\ 0\ n\ 0\ n\ l\ m\ 0\ 0$.
35. $k\ k\ n\ n\ n\ n\ l\ m\ 0\ n$.

The space 28 intersects V in a twisted cubic curve; 29 intersects V in five points; 30 intersects V in four points and contains a Σ-line tangent to V at one of them; 31 intersects V in four points but contains no tangent line at any of them; each of the others has just three points on V. Space 32 contains a Σ-line not in $n=0$ and not tangent to V; 33 contains one Σ-line tangent to V; 34 contains two Σ-lines tangent to V at two distinct points; 35 contains no Σ-line except the sides of the triangle determined by the three points of V.

Three-spaces with two points on V.

36. $k+n\ l\ n\ 0\ 0\ m\ l\ k+rm\ m\ 0$, x^3+rx-1 irreducible.
37. $k\ l\ 0\ -n\ n\ 0\ l\ m\ 0\ n$.
38. $k\ l\ 0\ 0\ n\ 0\ l\ m\ 0\ n$.
39. $k\ l\ 0\ n\ rn\ 0\ l\ m\ 0\ 0$.
40. $k\ l\ 0\ n\ 0\ n\ l\ m\ n\ 0$.
41. $k\ l\ 0\ n\ 0\ n\ l\ m\ 0\ 0$.
42. $k\ l\ n\ n\ n\ 0\ l\ m\ 0\ 0$.

Space 36 contains the τ-plane $n=0$ which has no point on V; the others contain τ-planes every one of which has at least one point on V. Spaces 37 and 38 have three τ-planes each; in 37 one of the τ-planes contains both points of V, in 38 two of the τ-planes contain both points of V. All of the planes on P_3 in 39 are τ-planes, and so is the plane $m=0$. Space 40 contains two τ-planes. Spaces 41 and 42 have pencils of τ-planes on the line joining the two points of V and in each the plane $m=0$ is a τ-plane. To distinguish between 41 and 42 an argument analogous to that given for planes 14 and 15 suffices. The points of X determine in each of spaces 41 and 42 a four-parameter system W of quadrics. In W there are two special pencils of quadrics determined by the two points on V. In 41 each quadric in either of the special pencils consists of a pair of planes; in 42 one of the pencils contains two quadrics each of which is a plane counted twice.

Three-spaces with one point on V.

43. $k\ l\ 0\ 0\ n\ m\ l\ k+rm\ m\ 0$, x^3+rx-1 irreducible.
44. $k\ l\ 0\ n\ m\ n\ l\ k\ 0\ 0$.
45. $k\ l\ n\ 0\ n\ m\ l\ k\ 0\ 0$.
46. $k\ l\ 0\ -n\ n\ m\ l\ k\ 0\ 0$.
47. $k\ l\ n\ 0\ 0\ m\ l\ k\ n\ 0$.
48. $k+n\ l\ m\ 0\ 0\ rn\ l\ k\ n\ 0$, x^3-x-r^2 irreducible.
49. $k\ l\ n\ 0\ 0\ m\ l+m\ k\ n\ 0$.
50. $k\ l\ 0\ -n\ n\ m\ l+m\ k\ 0\ 0$.
51. $k\ l\ n\ -n\ n\ 0\ l\ k\ 0\ m$.
52. $k\ l\ n\ n\ -n\ 2n\ l\ k\ 0\ m$.

The space 43 is tangent to V at P_4 which is on V; none of the others has this property. Spaces 44 and 45 each contain a plane tangent to V at O, the point on V; in 45 this tangent plane is a Σ-plane, in 44 it is not.

Spaces 46, 47, 48 intersect the space tangent to V at O in a line; 46 contains two τ-planes, 47 and 48 each contain only one τ-plane; in 47 the τ-plane passes through O, in 48 it does not. The space tangent to V at O intersects none of the other spaces anywhere except at O. Space 49 contains a single τ-plane, space 50 contains two τ-planes. Spaces 51 and 52 contain no τ-planes; each contains a special pencil of quadrics belonging to the four-parameter set W determined by the points of X. In both cases the quadrics of the special pencils are cones; in 51 three of the cones are degenerate, each consists of a pair of planes, in 52 only one is degenerate.

Three-spaces with no point on V.

53. $k\ l\ m\ 0\ m\ n\ l\ k+n\ n\ 0.$ 54. $k\ l\ 0\ 2n\ m+3n\ n\ l\ k\ 0\ m.$

Space 53 contains a τ-plane and a Σ-line; 54 contains neither. In 53 the τ-plane is $m = 0$ and the Σ-line $l = n = 0$. These special elements are made use of to get the canoncial form. There are no special lines or planes in 54. These facts are easily verified and are enough to distinguish between 53 and 54.

We mention another way in which 53 and 54 can be distinguished. In 54 the plane $n = 0$ is number 9 of the list of planes; every plane of 54 is of this type. Every plane of 53 except the τ-plane and the planes of the pencil on the Σ-line is of the same type. In either case we may take the plane $P_1P_2P_3$ to be $k\ l\ 0\ 0\ m\ 0\ l\ k\ 0\ m$ and $P_4 = 0\ 0\ a_3\ a_4\ a_5\ a_6\ a_7\ a_8\ a_9\ 0$. In either case the conditions for a single point on V lead to a polynomial of degree five in a single variable. In case 54 the polynomial is irreducible; in case 53 the polynomial is the product of a quadratic and a cubic, both irreducible in GF(p).

5. Some remarks about the groups. We make a few observations about the groups determined in the preceding pages. There are 85 groups: 1 of order p^{15}; 2 of order p^{14}; 6 of order p^{13}; 22 of order p^{12}; 54 of order p^{11}. Two groups of the same order have commutator subgroups of the same order; the index of the commutator subgroup is always p^5.

Let us change the notation to denote the commutators of pairs of generators by s_k instead of c_{ij} so that the commutator of x and y is $\Pi s_k^{a_k}$, where $1 \leqq k \leqq 10$. The group G, of order p^{15}, was defined by relations on the generators of which one was the requirement that there be no relations among the s_i's. For each of the other groups there will be relations among them.

Let S_i, $i = 0, 1, 2, 3$, be a subspace of dimension i of S. Let H be the subgroup of order p^{i+1} of C which is determined by S_i. Let G' and C' be the quotient groups G/H and C/H respectively. Then G' will be defined by the relations which define G together with $i+1$ relations which put equal to the identity commutators which correspond to $i+1$ independent points of S_i; ordinarily the simplest relations are obtained if commutators corresponding to the vertices of the frame of reference in S_i are put equal to the identity. For example, if S_2 is plane 9, the simplest additional relations required to define G' are: $s_1s_8 = 1$, $s_2s_7 = 1$, $s_5s_{10} = 1$.

We have given close attention to the subgroups of G determined by subspaces of S; it will be useful to look more closely at some subgroups of G determined by subspaces of X. A point in X determines a subgroup α of order p in the abelian quotient group G/C and determines an abelian subgroup of order p^{11} in G. Two abelian subgroups of order p^{11} of G are conjugate under the group of isomorphisms of G, but every abelian subgroup of order p^{11} is self-conjugate in G. A line in X determines a subgroup of order p^{12} of G and this subgroup is non-abelian with commutator subgroup of order p. This subgroup is also self-conjugate in G but is conjugate to any other such subgroup under the group of isomorphism of G. Such a subgroup is $\{C, U_1, U_2\} = \Gamma$. Γ is the direct product of the metabelian group $\{U_1, U_2\}$ which contains the commutator s_1 and the abelian group $\{s_2, s_3, \cdots, s_{10}\}$. Γ is also the direct product of the abelian group just given and any group $\{cU_1, c'U_2\}$ where c and c' are arbitrary elements of C. The group $\{s_2, s_3, \cdots, s_{10}\}$ is not, however, uniquely determined by the line in X since it may be replaced by any subgroup of order p^9 of C which does not contain s_1. A line in X therefore determines a unique subgroup of G of order p^{12}, and determines a set of metabelian subgroups of order p^3 all of which have for a commutator subgroup the cyclic group determined by the point in S which is the image of the line in X. A plane in X determines a subgroup of order p^{13} of G, and a set of metabelian subgroups of order p^6 all of which have the same commutator subgroup of order p^3. A three-space in X determines a unique subgroup of order p^{14}, and a set of metabelian subgroups each generated by four elements, each of order p^{10}, and all with the same commutator subgroup of order p^6.

The groups G' of order p^{14}, determined by the two types of point in S, differ in the fact that the first one contains an abelian subgroup $\{C', U_1, U_2\}$ of order p^{11} and the second contains no abelian subgroup of order p^{11}. The fact that the point of the second type determines a unique three-space R in X is reflected in the fact that the corresponding G' contains a unique subgroup of order p^{13} which has a commutator subgroup of order p^5.

The lines of S determines six types of group of order p^{13}, each with commutator subgroup of order p^8. In G' which corresponds to line 1 of the list of lines there are $p+1$ abelian subgroups $\{C', U_1, U_2^k, U_3^l\}$ of order p^{10} all contained in the non-abelian subgroup $\{C', U_1, U_2, U_3\}$ of order p^{11}. The second line of the list gives a group G' which contains two abelian subgroups of order p^{10}, viz., $\{C', U_1, U_2\}$ and $\{C', U_3, U_4\}$; both abelian subgroups are contained in the group $\{C', U_1, U_2, U_3, U_4\}$ of order p^{12} whose commutator subgroup has order p^4; this last is the only subgroup of order p^{12} with commutator subgroup of order less than p^5. Lines 3 and 4 of the list determine groups G' which have respectively one and no abelian subgroup of order p^{10}. Both have subgroups of order p^{12} with commutator subgroups of order p^4. Lines 5 and 6 determine groups G' which also have respectively one and no abelian subgroups of order p^{10}. Neither contains a subgroup of order p^{12} with commutator subgroup of order as small as p^4.

The group G' determined by line number 6 in the list is defined by the (additional) relations: $s_1 s_8 = 1$, $s_2 s_7 = 1$. Let us denote by L the line $k\,l\,0\,0.0\,0\,l\,k\,0\,0$. Every point of L determines a subgroup Γ of G with commutator subgroup β; Γ is of order p^{14} and β is of order p^6. Corresponding to these groups are subgroups Γ' and β' of order p^{12} and p^5 in G'. The line L determines the point $M = 0\,0\,0\,0\,1\,0\,0\,0\,0\,0$ on V. The line m is in the three-space R determined in X by any point P of L. Hence, the non-abelian group $\{C', U_2, U_3\}$ of order p^{10} is in every one of the $p+1$ special subgroups Γ' of G', which are special because they are the only subgroups of order p^{12} with commutator subgroups of order p^5.

The above examples are meant to suggest the complexity of the relations among subgroups of the groups G', but also to suggest that these complex relations are not, in these groups at least, unmanageable.

UNIVERSITY OF ILLINOIS.

HYPERGEODESICS AND DUAL HYPERGEODESICS.*

By T. K. PAN.

1. **Introduction.** Dual union curves on a surface were first defined and studied by P. Sperry [4, p. 222] and investigated later by G. M. Green [2, pp. 140-144]. In this paper a system of curves on a surface called dual hypergeodesics is defined, which includes a system of dual union curves as a special case, its intrinsic differential equation is found, most of the theorems for dual union curves are generalized and some properties of hypergeodesics are studied.

The notation of Eisenhart [1] will be used for the most part except that $\Gamma^{\alpha}{}_{\beta\gamma}$ will be employed for Christoffel symbols of the second kind.

2. **The extended relation *R*.** Consider a real proper analytic surface S defined with reference to a rectangular cartesian coordinate system by

$$x^i = x^i(u^1, u^2) \qquad (i = 1, 2, 3). \tag{2.1}$$

The coordinate curves are assumed to form the asymptotic net. The tangents to one family of the net constructed at the points of a curve of the other family generate a non-developable ruled surface. There are two such surfaces R_1, R_2 associated with $P(x^i)$. Any two planes different from the tangent plane of S at P and passing through the asymptotic tangents at P will be tangent to R_1, R_2 at two points P_1, P_2 respectively. The intersecting line L' of these two planes passes through P but does not lie on the tangent plane of S at P; and the line joining P_1, P_2 lies on the tangent plane of S at P but does not pass through P. These two lines L', L are said to be in the relation R with respect to the asymptotic net [3, p. 81]. As P varies over S, they generate two congruences Γ', Γ respectively.

Let the direction cosines of L' be defined by

$$\lambda^i = p^{\alpha}\partial x^i/\partial u^{\alpha} + qX^i \tag{2.2}$$

where X^i are the direction cosines of the normal to S at P, and where p^{α}, q

---

* Received May 16, 1950.

are functions of u^α, q being assumed to be always positive. The coordinates of the two points P_α are found to be

$$\bar{x}^i{}_\alpha = x^i - q/(q\Gamma^\beta{}_{12} - d_{12}p^\beta)\partial x^i/\partial u^\alpha, \qquad \alpha \neq \beta. \tag{2.3}$$

In terms of the local coordinates y_1, y_2 which are defined as the coordinates of the terminal point of the vector

$$y_1\left(\frac{1}{\sqrt{g_{11}}}\frac{\partial x^i}{\partial u^1}\right) + y_2\left(\frac{1}{\sqrt{g_{22}}}\frac{\partial x^i}{\partial u^2}\right)$$

when it is localized at P, the equation of L is readily found to be

$$\sqrt{g_{22}}(q\Gamma^2{}_{12} - d_{12}p^2)y_1 + \sqrt{g_{11}}(q\Gamma^1{}_{12} - d_{12}p^1)y_2 + q\sqrt{(g_{11}g_{22})} = 0. \tag{2.4}$$

When the asymptotic tangents at P are considered as the lines L', the lines L in the relation R to them with respect to the asymptotic net are defined to be given by (2.4). We call the relation R with the addition of these pairs of lines *the extended relation R*. Geometrically, an asymptotic tangent at P on S is in the extended relation R to itself with respect to the asymptotic net.

3. Definition and equation of dual hypergeodesics. The integral curves of a differential equation of the form

$$d^2u^2/du^{12} = A + Bdu^2/du^1 + C(du^2/du^1)^2 + D(du^2/du^1)^3 \tag{3.1}$$

in which A, B, C, D are analytic functions of u^α, are called hypergeodesics on S. They form a two-parameter family. The envelope of their osculating planes through P on S is a cone with vertex at P and has the tangent plane of S at P for bitangent plane, the lines of contact being the asymptotic tangents of S at P. This cone which is called the osc-cone at P of the family is ordinarily of class three and order four [3, p. 101]. Except the two asymptotic tangents all generators of the osc-cone at P are lines L'. Hence we can always find the line L which is in the extended relation R to each generator of the osc-cone at P of a family of hypergeodesics.

A dual hypergeodesic relative to a family of hypergeodesics will be defined as a curve on S such that the ray point at each point P of the curve lies on the envelope of the lines L which are in the extended relation R to the generators of the osc-cone at P of the family with respect to the asymptotic net of curves on S.

Let $u^2 = u^2(u^1)$ be a hypergeodesic C through P of the family (3.1) such that its direction at P is given by $\lambda = du^2/du^1$. The osculating plane of C at $P(x^i)$ is found by the aid of the equations of Gauss to be

$$(3.2) \quad e_{ijk}(\bar{x}^i - x^i)\left(\frac{\partial x^j}{\partial u^1} + \frac{\partial x^j}{\partial u^2}\lambda\right)\left[\frac{\partial x^k}{\partial u^2}(A + B\lambda + C\lambda^2 + D\lambda^3) + \frac{\partial x^k}{\partial u^\gamma}(\Gamma^\gamma{}_{11} + 2\Gamma^\gamma{}_{12}\lambda + \Gamma^\gamma{}_{22}\lambda^2) + 2d_{12}\lambda X^k\right] = 0.$$

Equation (3. 2) and the equation

$$\begin{aligned}(3.3) \quad e_{ijk}(\bar{x}^i - x^i)\Big[&\frac{\partial x^j}{\partial u^1}\frac{\partial x^k}{\partial u^2}(B + 2C\lambda + 3D\lambda^2) \\ &+ 2\frac{\partial x^j}{\partial u^1}\frac{\partial x^k}{\partial u^\gamma}(\Gamma^\gamma{}_{12} + \Gamma^\gamma{}_{22}\lambda) + 2d_{12}X^k\frac{\partial x^j}{\partial u^1} \\ &+ \frac{\partial x^j}{\partial u^2}\frac{\partial x^k}{\partial u^2}(A + 2B\lambda + 3C\lambda^2 + 4D\lambda^3) \\ &+ \frac{\partial x^j}{\partial u^2}\frac{\partial x^k}{\partial u^\gamma}(\Gamma^\gamma{}_{11} + 4\Gamma^\gamma{}_{12}\lambda + 3\Gamma^\gamma{}_{22}\lambda^2) + 4d_{12}\lambda X^k\frac{\partial x^j}{\partial u^2}\Big] = 0\end{aligned}$$

define two planes intersecting in a generator L' of the osc-cone of the family (3. 1) corresponding to the hypergeodesics C in the direction λ at P. The direction numbers of the normal to each of these planes are

$$\begin{aligned}(3.4) \quad a_m = \sqrt{g}X^m[&(A + \Gamma^2{}_{11}) + (B + 2\Gamma^2{}_{12} - \Gamma^1{}_{11})\lambda \\ &+ (C + \Gamma^2{}_{22} - 2\Gamma^1{}_{12})\lambda^2 + (D - \Gamma^1{}_{22})\lambda^3] \\ &+ 2d_{12}e_{mjk}\frac{\partial x^j}{\partial u^1}X^k\lambda + 2d_{12}e_{mjk}\frac{\partial x^j}{\partial u^2}X^k\lambda^2\end{aligned}$$

$$\begin{aligned}(3.5) \quad b_n = \sqrt{g}X^n[&(B + 2\Gamma^2{}_{12} - \Gamma^1{}_{11}) \\ &+ 2(C + \Gamma^2{}_{22} - 2\Gamma^1{}_{12})\lambda + 3(D - \Gamma^1{}_{22})\lambda^2] \\ &+ 2d_{12}e_{njk}\frac{\partial x^j}{\partial u^1}X^k + 4d_{12}e_{njk}\frac{\partial x^j}{\partial u^2}X^k\lambda.\end{aligned}$$

Hence the generator L' has the direction numbers

$$(3.6) \qquad c^g = e^{gmn}a_m b_n$$

which, by the aid of the formula [1, p. 8, Ex. 9]

$$(3.7) \qquad e^{ijh}e_{klh} = \delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk}$$

and after discarding of irrelevant non-zero multipliers, reduces to

$$(3.8) \qquad c^i = P^\alpha \partial x^i/\partial u^\alpha + QX^i$$

where

$$\begin{aligned} &P^1 = (A + \Gamma^2{}_{11}) - (C + \Gamma^2{}_{22} - 2\Gamma^1{}_{12})\lambda^2 - 2(D - \Gamma^1{}_{22})\lambda^3, \\ (3.9)\qquad &P^2 = 2(A + \Gamma^2{}_{11})\lambda + (B + 2\Gamma^2{}_{12} - \Gamma^1{}_{11})\lambda^2 - (D - \Gamma^1{}_{22})\lambda^4, \\ &Q = 2d_{12}\lambda^2. \end{aligned}$$

Substituting (3.9) into (2.4) and arranging the terms as a polynomial in λ we obtain the following equation of L, which is in the extended relation R to the generator L'

$$(3.10)\qquad a\lambda^4 + b\lambda^3 + c\lambda^2 + d\lambda + e = 0,$$

where the coefficients are defined by

$$\begin{aligned} &a = \sqrt{g_{22}}(D - \Gamma^1{}_{22})y_1, \qquad b = 2\sqrt{g_{11}}(D - \Gamma^1{}_{22})y_2, \\ (3.11)\qquad &c = -\sqrt{g_{22}}(B - \Gamma^1{}_{11})y_1 + \sqrt{g_{11}}(C + \Gamma^2{}_{22})y_2 + 2\sqrt{(g_{11}g_{22})}, \\ &d = -2\sqrt{g_{22}}(A + \Gamma^2{}_{11})y_1, \qquad e = -\sqrt{g_{11}}(A + \Gamma^2{}_{11})y_2. \end{aligned}$$

As λ varies, the line L (3.10) generates an envelope in the tangent plane of S at P, whose equation is found after considerable simplification to be

$$(3.12)\qquad ad^2 + eb^2 + bcd = 0$$

which from (3.11) becomes

$$\begin{aligned} (3.13)\qquad &(A + \Gamma^2{}_{11})(y_1/\sqrt{g_{11}})^3 + (-D + \Gamma^1{}_{22})(y_2/\sqrt{g_{22}})^3 \\ &- y_1y_2/\sqrt{(g_{11}g_{22})}[(-B + \Gamma^1{}_{11})y_1/\sqrt{g_{11}} + (C + \Gamma^2{}_{22})y_2\sqrt{g_{22}} + 2] = 0. \end{aligned}$$

This envelope has a node at P, the nodal tangents being the asymptotic tangents of the surface at P. The curve is ordinarily of the third order and will be called *the reciprocal cubic* at P of the family (3.1).

The ray point at P of a curve $u^\alpha = u^\alpha(s)$ on S is by definition the intersection of the planes

$$\begin{aligned} &\Sigma(\bar{x}^i - x^i)X^i = 0, \qquad \Sigma(\bar{x}^i - x^i)u'^\alpha \partial X^i/\partial u^\alpha = 0, \\ (3.14)\qquad & \\ &\Sigma(\bar{x}^i - x^i)\left(\frac{\partial^2 X^i}{\partial u^\alpha \partial u^\beta}u'^\alpha u'^\beta + \frac{\partial X^i}{\partial u^\alpha}u''^\alpha\right) = \frac{\partial X^i}{\partial u^\alpha}\frac{\partial x^i}{\partial u^\beta}u'^\alpha u'^\beta. \end{aligned}$$

Solving simultaneously, C. E. Springer found its local coordinates [5, p. 904] to be

$$(3.15)\qquad y_\delta = \sqrt{g_{\delta\delta}} \cdot dd_{\alpha\beta}u'^\alpha u'^\beta \frac{e_{\gamma\sigma}d^{\sigma\delta}u'^\gamma}{e^{\sigma\tau}(d_{\gamma\sigma}d_{\mu\tau}\bar{\Gamma}^\mu{}_{\alpha\beta}u'^\alpha u'^\beta u'^\gamma + d_{\alpha\sigma}d_{\beta\tau}u'^\alpha u''^\beta)}$$

where δ is not summed and where $\bar{\Gamma}^\mu{}_{\alpha\beta} = d^{\nu\gamma}\partial d_{\alpha\nu}/\partial u^\beta - d^{\nu\gamma}d_{\alpha\mu}\Gamma^\mu{}_{\nu\beta}$. Substituting (3.15) into (3.13), we find that the dual hypergeodesics relative to the family (3.1) with respect to the asymptotic net are defined by the equation

$$(3.16)\quad d^2u^2/du^{1^2} = -A + Bdu^2/du^1 + C(du^2/du^1)^2 - D(du^2/du^1)^3.$$

When (3.1) is a family of union curves, (3.16) evidently reduces to the equation of dual union curves which include dual geodesics [5, p. 904] as a special case.

4. Reciprocal characteristics.

Since equation (3.16) is of the same form as (3.1), it defines a two parameter family of hypergeodesics. When (3.1) is given, (3.16) is uniquely determined. Hence

THEOREM 4.1. *The dual hypergeodesics relative to a given two-parameter family of hypergeodesics on a surface are hypergeodesics. They form a two parameter family called the dual of the given family. There exists one and only one family of hypergeodesics which is the dual of a given two-parameter family of hypergeodesics.*

Equation (3.16) differs from equation (3.1) in having the signs of A and D changed. It is evident that each represents the dual of the family of hypergeodesics defined by the other. Hence

THEOREM 4.2. *The dual relation between any two two-parameter families of hypergeodesics on a surface is reciprocal.*

A family of hypergeodesics and the dual of the family will be called *reciprocal families.* Thus (3.1) and (3.16) represent reciprocal families of hypergeodesics. In particular the union curves of a congruence Γ' on a surface are the dual hypergeodesics relative to the family of dual union curves of the congruence Γ, and therefore the geodesics on a surface are the dual hypergeodesics relative to the family of dual geodesics.

An asymptotic line on a surface is always a hypergeodesic on the surface, since the coefficients in (3.1) are arbitrary and $A = 0$ gives $u^2 =$ const. as a solution of (3.1). But from (3.16) $A = 0$ is also the condition that $u^2 =$ const. is a dual hypergeodesic. Hence an asymptotic line is at the same time a hypergeodesic of two reciprocal families of hypergeodesics.

Any straight line on a surface is an asymptotic line and also a union curve on the surface, since a geodesic is a special union curve. The osculating planes of a curved asymptotic line on a surface are tangent to the surface along the curve and consequently can not pass through any line L' of a congruence Γ'. Hence there is no curved asymptotic line on a surface, which is a union curve on the surface. Thus straight lines on a surface are the only union curves which are asymptotic lines.

If an asymptotic line is a hypergeodesic of every two-parameter family of hypergeodesics (3.1) for which $A = 0$ or $D = 0$, then it can be a dual union curve and in that case it is a straight line. Hence

THEOREM 4.3. *If an asymptotic line is a hypergeodesic of every family* (3.1) *for which* $A = 0$ *or* $D = 0$, *it is a straight line.*

When a family of hypergeodesics is identical with its reciprocal, we say the family is *self dual.* Equations (3.1), (3.16) become identical if and only if

$$A \equiv 0, \qquad D \equiv 0, \tag{4.1}$$

which by Theorem 4.3 imply that the asymptotic lines on the surface are straight lines. Consequently the surface is a quadric. Hence

THEOREM 4.4. *Every two-parameter family of hypergeodesics on a surface is self dual if and only if the surface is a quadric.*

When a two-parameter family of hypergeodesics is not self dual, there may be a particular hypergeodesic of the family coincident with a hypergeodesic of its reciprocal family. From equations (3.1), (3.16) such a curve is a solution of the two equations

$$A + D(du^2/du^1)^3 = 0, \tag{4.2}$$

$$d^2u^2/du^{12} = Bdu^2/du^1 + C(du^2/du^1)^2. \tag{4.3}$$

The curves defined by (4.2), their directions and the tangents in these directions will be called respectively *the generalized curves of Segre, the gneralized directions of Segre,* and *the generalized tangents of Segre, associated with the family* (3.1) *or* (3.16).

The solutions of equation (4.2) do not necessarily satisfy equation (4.3). There may exist no solution or one solution of (4.2) which satisfies (4.3). In case $AD \neq 0$, when two solutions of (4.2) satisfy (4.3), it can be shown easily that the third solution will also satisfy it. Hence

THEOREM 4.5. *The hypergeodesics common to two reciprocal families of hypergeodesics are the generalized curves of Segre associated with either one of these families. Conversely, two reciprocal families of hypergeodesics do not necessarily have hypergeodesics in common except reciprocal families on a quadric surface or those families on a general surface for which*

$A = D = 0$, which are always identical. On a non-ruled surface those reciprocal families for which $AD \neq 0$, may have no hypergeodesics in common or may have one or three one-parameter families of hypergeodesics in common.

5. Conjugate nets. If the family (3.1) contains a conjugate net represented by

$$(5.1) \qquad (du^2)^2 - \lambda^2 (du^1)^2 = 0$$

where λ is a scalar function of u^a, we have

$$(5.2) \qquad \phi_{u^2} = 2A + 2C\phi, \qquad \phi_{u^1} = 2B\phi + 2D\phi^2,$$

where A, B, C, D are supposed to be known functions and $\phi = \lambda^2$ is to be determined. A necessary and sufficient condition for a common solution of (5.2) is that

$$(5.3) \qquad \partial(A + C\phi)/\partial u^1 = \partial(B\phi + D\phi^2)/\partial u^2$$

be satisfied identically by such a solution. This condition after differentiation becomes

$$(5.4) \qquad (D_{u^2} + 2CD)\phi^2 + (B_{u^2} - C_{u^1} + 4AD)\phi + (2AB - A_{u^1}) = 0$$

which is either satisfied identically in ϕ at each point or not. In any case we can show that A, B, C, D can not all be arbitrary and obtain the following result.

THEOREM 5.1. *A two-parameter family of hypergeodesics on a surface does not in general contain a conjugate net, but a family can be chosen in infinitely many ways so that its hypergeodesics contain one conjugate net or a one-parameter family of conjugate nets.*

A similar result can be obtained for the pairing of the hypergeodesics of a family (3.1) and the dual hypergeodesics of its reciprocal family (3.16) to form a conjugate net.

The condition that (3.16) will contain the associate conjugate net of the net (5.1), defined by

$$(5.5) \qquad (du^2)^2 + \lambda^2 (du^1)^2 = 0,$$

is found exactly the same as (5.2). Hence

THEOREM 5.2. *If there exists on a surface a conjugate net formed of hypergeodesics of a two-parameter family of hypergeodesics, then the associate conjugate net of it consists of the dual hypergeodesics of the reciprocal family; and conversely.*

6. Associated cones and plane curves. In the homogeneus local cartesian coordinates v_1, v_2, v_3, v_4 of the plane $v_1y_1 + v_2y_2 + v_3y_3 + v_4 = 0$, the equation of the cusp axis of the osc-cone of (3.1) at P are found to be

$$(6.1) \quad -\sqrt{g_{11}}(C + \Gamma^2_{22} - 2\Gamma^1_{12})v_1 + \sqrt{g_{22}}(B + 2\Gamma^2_{12} - \Gamma^1_{11})v_2 + 2d_{12}v_3 = 0, \qquad v_4 = 0,$$

and its three cuspidal tangent planes are found to intersect the tangent plane of S at P in the three tangents

$$(6.2) \quad (\sqrt{g_{11}})^3(-D + \Gamma^1_{22})v^3_1 + (\sqrt{g_{22}})^3(A + \Gamma^2_{11})v^3_2 = 0, \qquad v_4 = 0.$$

These three tangents will be the same as those determined by the reciprocal family (3.16) if and only if

$$(6.3) \qquad A = \rho\Gamma^2_{11}, \qquad D = -\rho\Gamma^1_{22},$$

where $\rho \neq 1$ is either a constant or a function of u^α. In this case the directions of the three tangents become those of Segre, since $y_1/\sqrt{g_{11}} = dw^1$, $y_2/\sqrt{g_{22}} = dw^2$. Hence

THEOREM 6.1. *The osc-cones of two reciprocal families of hypergeodesics at a point P always have a common cusp axis. They have the same cuspidal tangent planes if and only if the latter intersect the tangent plane of S at P in its Segre tangents.*

The locus of the ray points of all the hypergeodesics through P of a family (3.1) is a plane cubic curve called the ray-point cubic at P of the family. Its equation in local coordinates on the tangent plane is found to be

$$(6.4) \quad (-A + \Gamma^2_{11})(y_1/\sqrt{g_{11}})^3 + (D + \Gamma^1_{22})(y_2/\sqrt{g_{22}})^3 - y_1y_2/\sqrt{(g_{11}g_{22})}\,[(-B + \Gamma^1_{11})y_1/\sqrt{g_{11}} + (C + \Gamma^2_{22})y_2/\sqrt{g_{22}} + 2] = 0.$$

A comparison of (6.4) and (3.13) and some calculation yield the results:

THEOREM 6.2. *The reciprocal cubic of a family of hypergeodesics is the ray-point cubic of its reciprocal family. The ray-point cubics or the reciprocal cubics at a point of two reciprocal families of hypergeodesics have a common flex ray. They possess a common set of points of inflection if and only if the three points are on the Darboux tangents.*

THEOREM 6.3. *The osc-cones of two reciprocal families of hypergeodesics possess three common cuspidal tangent planes if and only if their ray-point cubics or reciprocal cubics possess three common points of inflection.*

Since two reciprocal families have the same cusp axis, they have the same cusp axis curves, which form a conjugate net if and only if $B_{u^2} + C_{u^1} = 0$. Since a self dual family consists of a pencil of conjugate nets if and only if $B_{u^2} - C_{u^1} = 0$, we have

THEOREM 6. 4. *The cusp axis curves of a self dual family of hypergeodesics which contains a pencil of conjugate nets form a conjugate net if and only if B is a function of u^1 alone and C is a function of u^2 alone.*

In general the osc-cones at a point of two reciprocal families of hypergeodesics have sixteen generators in common, four of which coincide with the two asymptotic tangents at the point. The remaining twelve common generators are in general distinct. For if two of the non-asymptotic common generators coincide with L', the two osc-cones are tangent to the same plane along L' and this requires $A - 2D\lambda^3 = 0$, $2A - D\lambda^3 = 0$, that is $A = 0$, $D = 0$, since $\lambda \neq 0$ or ∞. Hence

THEOREM 6. 5. *The osc-cones of two reciprocal families do not have double contact along any non-asymptotic common generator unless the family is self dual or the surface is a quadric.*

The direction λ at which the osc-cones of two reciprocal families have a common tangent plane along different generators is found to satisfy

$$A + D\lambda^3 = 0. \tag{6. 5}$$

Since λ is non-asymptotic, $A = 0$ and $D \neq 0$ or $A \neq 0$ and $D = 0$ can not satisfy (6. 5). In case $A = D = 0$, the osc-cones are identical. Hence

THEOREM 6. 6. *On a non-ruled surface the osc-cones at P of two reciprocal families for which $AD \neq 0$ have, besides the plane tangent to S at P, three common tangent planes in the generalized directions of Segre associated with either of the families.*

The osc-cones at P of two reciprocal families for which $A = 0$, $D \neq 0$ or $A \neq 0$, $D = 0$ on a non-ruled surface and those for which $A \neq 0$ or $D \neq 0$ on a ruled surface have only one common tangent plane—the tangent plane to the surface at P. This plane corresponds to an asymptotic direction which is now a generalized direction of Segre associated with the respective family under consideration. The osc-cones at P of two reciprocal families for which $A = 0$, $D = 0$ on a non-ruled surface and those on a quadric are identical and have therefore infinitely many common tangent planes corresponding to the infinitely many directions of Segre associated with the respective family under consideration. With Theorem 6. 6 we may sum up these results in the following theorem:

THEOREM 6. 7. *The generalized directions of Segre associated with a family of hypergeodesics are the directions in which the osc-cones of the family and its reciprocal family have common tangent planes.*

The following two theorems can be easily verified:

THEOREM 6. 8. *The axis cones at a point of two reciprocal families of hypergeodesics have double contact along the asymptotic tangents on two planes which intersect in their common cusp axis at the point.*

THEOREM 6. 9. *The ray conics at a point of two reciprocal families have double contact on the asymptotic tangents at the points where the flex ray at P cuts them.*

7. Torsal curves. Let C defined by $u^2 = u^2(u^1)$ be a hypergeodesic C of (3. 1) such that the direction of the curve is determined by $\lambda = du^2/du^1$. We shall determine λ so that C is a curve such that the generators corresponding to λ of the osc-cone of (3. 1) at points of C form a developable surface and we shall call the curve *a torsal curve of the family* (3. 1).

If C is an asymptotic line, its corresponding generators on the osc-cone are the asymptotic tangents generating a developable of which the asymptotic line is both an edge of regression and a torsal curve. Consequently an asymptotic hypergeodesic of a family (3. 1) is always a torsal curve of the family and it is a plane curve if and only if it is a straight line. In the following discussion we shall exclude the asymptotic lines by assuming $\lambda \neq 0$ or ∞.

Looked upon as a curve in space the tangent at a point P of C has its direction numbers given by

$$(7.1) \qquad dx^i/du^1 = \partial x^i/\partial u^\alpha \cdot du^\alpha/du^1 = \partial x^i/\partial u^1 + \lambda \partial x^i/\partial u^2.$$

The direction numbers of the generator of the osc-cone corresponding to C at P are given by (3. 8). Let M be any point different from P on the generator. Then M may be defined by $\xi^i = x^i + \rho c^i$, where ρ is an arbitrary but fixed constant different from zero. The tangent to the curve traced by M as P moves on C has its direction numbers given by

$$(7.2) \qquad \rho^{-1} d\xi^i/du^1 = \bar{P}^\alpha \partial x^i/\partial u^\alpha + \bar{Q} X^i,$$

where

$$
\begin{aligned}
\bar{P}^1 &= 1/\rho + P^1\Gamma^1{}_{11} + \partial P^1/\partial u^1 + P^2\Gamma^1{}_{12} - g^{12}d_{12}R + P^1\Gamma^1{}_{12}\lambda \\
&\qquad + \lambda\partial P^1/\partial u^2 + P^2\Gamma^1{}_{22}\lambda - g^{11}d_{12}R\lambda, \\
\bar{P}^2 &= \lambda/\rho + P^1\Gamma^2{}_{11} + P^2\Gamma^2{}_{12} + \partial P^2/\partial u^1 - g^{22}d_{12}R \\
&\qquad + P^1\Gamma^2{}_{12}\lambda + P^2\Gamma^2{}_{22}\lambda + \lambda\partial P^2/\partial u^2 - g^{12}d_{12}R\lambda, \\
\bar{Q} &= P^2d_{12} + \partial R/\partial u^1 + P^2d_{12}\lambda + \lambda\partial R/\partial u^2,
\end{aligned}
\tag{7.3}
$$

which are obtained after simplification by the aid of the Gauss and the Weingarten equations.

The generator of the osc-cone will generate a developable as P moves on C if and only if the tree lines—the tangent to C, the corresponding generator of the osc-cone and the tangent to the curve traced by M—are coplanar, that is

$$
e_{ijk}(\partial x^i/\partial u^\alpha \cdot du^\alpha/du^1)\,(P^\beta\partial x^j/\partial u^\beta + QX^j)\,(\bar{P}^\gamma\partial x^k/\partial u^\gamma + \bar{Q}X^k) = 0.
\tag{7.4}
$$

This reduces to

$$
t_0 + t_1\lambda + t_2\lambda^2 + t_3\lambda^3 + t_4\lambda^4 + t_5\lambda^5 + t_6\lambda^6 = 0,
\tag{7.5}
$$

where a factor λ^2 is omitted and where the coefficients are defined by

$$
\begin{aligned}
t_0 &= 3A^2 + 4A\Gamma^2{}_{11} + (\Gamma^2{}_{11})^2, \\
t_1 &= 4AB + 2A\Gamma^2{}_{12} - 2A\Gamma^1{}_{11} + 4B\Gamma^2{}_{11} + 2\Gamma^2{}_{11}\Gamma^2{}_{12} - 2\Gamma^2{}_{11}\Gamma^1{}_{11} \\
&\quad + 2(A + \Gamma^2{}_{11})(\log d_{12})_{u^1} - 2(\Gamma^2{}_{11})_{u^1} - 2A_{u^1}, \\
t_2 &= 2AC - 2A\Gamma^2{}_{22} + 2A\Gamma^1{}_{12} + 4C\Gamma^2{}_{11} - 2\Gamma^2{}_{11}\Gamma^1{}_{12} + B^2 + (\Gamma^1{}_{11})^2 \\
&\quad + 2B\Gamma^2{}_{12} - 2B\Gamma^1{}_{11} - 2\Gamma^2{}_{12}\Gamma^1{}_{11} + 2(B + 2\Gamma^2{}_{12} - \Gamma^1{}_{11})(\log d_{12})_{u^1} \\
&\quad + 2(A + \Gamma^2{}_{11})(\log d_{12})_{u^2} - 2B_{u^1} - 4(\Gamma^2{}_{12})_{u^1} + 2(\Gamma^1{}_{11})_{u^1} \\
&\quad + 4g^{22}d^2{}_{12} - 2A_{u^2} - 2(\Gamma^2{}_{11})_{u^2}, \\
t_3 &= 2(C + \Gamma^2{}_{22} - 2\Gamma^1{}_{12})(\log d_{12})_{u^1} + 2(B + 2\Gamma^2{}_{12} - \Gamma^1{}_{11})(\log d_{12})_{u^2} \\
&\quad + 4D\Gamma^2{}_{11} + 2C\Gamma^2{}_{12} - 2B\Gamma^2{}_{22} - 2\Gamma^2{}_{12}\Gamma^2{}_{22} - 2B_{u^2} - 4(\Gamma^2{}_{12})_{u^2} \\
&\quad + 2(\Gamma^1{}_{11})_{u^2} - 2C\Gamma^1{}_{11} + 2\Gamma^1{}_{12}\Gamma^1{}_{11} - 2C_{u^1} - 2(\Gamma^2{}_{22})_{u_1} \\
&\quad + 4(\Gamma^1{}_{12})_{u^1} + 2B\Gamma^1{}_{12} + 4A\Gamma^1{}_{22}, \\
t_4 &= -2BD - 2D\Gamma^1{}_{11} + 2D\Gamma^2{}_{12} + 4B\Gamma^1{}_{22} + 2\Gamma^2{}_{12}\Gamma^1{}_{22} - C^2 \\
&\quad - (\Gamma^2{}_{22})^2 + 2C\Gamma^1{}_{12} - 2C\Gamma^2{}_{22} + 2\Gamma^2{}_{22}\Gamma^1{}_{12} + 2(D - \Gamma^1{}_{22})(\log d_{12})_{u^1} \\
&\quad + 2(C + \Gamma^2{}_{22} - 2\Gamma^1{}_{12})(\log d_{12})_{u^2} - 2C_{u^2} - 2(\Gamma^2{}_{22})_{\mathrm{u}^2} \\
&\quad + 4(\Gamma^1{}_{12})_{u^2} - 4g^{11}d^2{}_{12} - 2D_{u^1} + 2(\Gamma^1{}_{22})_{u^1}, \\
t_5 &= -4DC + 2D\Gamma^1{}_{12} - 2D\Gamma^2{}_{22} + 4C\Gamma^1{}_{22} + 2\Gamma^1{}_{22}\Gamma^2{}_{22} - 2\Gamma^1{}_{22}\Gamma^1{}_{12} \\
&\quad + 2(D - \Gamma^1{}_{22})(\log d_{12})_{u^2} + 2(\Gamma^1{}_{22})_{u^2} - 2D_{u^2}, \\
t_6 &= -3D^2 + 4D\Gamma^1{}_{22} - (\Gamma^1{}_{22})^2.
\end{aligned}
\tag{7.6}
$$

In the space in which the surface S: (2.1) is imbedded, the curve C: $u^2 = u^2(u^1)$ has equations $x^i = x^i(u^1, u^2(u^1)) = f^i(u^1)$. A necessary and sufficient condition that C be a plane curve is

$$e_{ijk} dx^i/du^1 \cdot d^2x^j/du^{12} \cdot d^3x^k/du^{13} = 0. \tag{7.7}$$

When we substitute in (7.7) the following values

$$\begin{aligned}
dx^i/du^1 &= \partial x^i/\partial u^1 + \lambda \partial x^i/\partial u^2, \\
d^2x^j/du^{12} &= p_1 \partial x^j/\partial u^1 + p_2 \partial x^j/\partial u^2 + rX^j, \\
d^3x^k/du^{13} &= \partial x^k/\partial u^1 [p_1\Gamma^1{}_{11} + \partial p_1/\partial u^1 + p_2\Gamma^1{}_{12} - 2d^2{}_{12}g^{12}\lambda' + p_1\Gamma^1{}_{12}\lambda \\
&\qquad + p_2\Gamma^1{}_{22}\lambda - 2d^2{}_{12}g^{11}\lambda^2 + \lambda\partial p_1/\partial u^2] \\
&\quad + \partial x^k/\partial u^2 [p_1\Gamma^2{}_{11} + p_2\Gamma^2{}_{12} + \partial p_2/\partial u^1 - 2d^2{}_{12}g^{22}\lambda \\
&\qquad + p_1\Gamma^2{}_{12}\lambda + p_2\Gamma^2{}_{22}\lambda + \lambda\partial p_2/\partial u^2 - 2d^2{}_{12}g^{12}\lambda^2] \\
&\quad + X^k[d_{12}p_2 + \partial r/\partial w^1 + p_1 d_{12}\lambda + \lambda \partial r/\partial u^2]
\end{aligned} \tag{7.8}$$

where $p_1 = \Gamma^1{}_{11} + 2\Gamma^1{}_{12}\lambda + \Gamma^1{}_{22}\lambda^2$, $p_2 = \Gamma^2{}_{11} + 2\Gamma^2{}_{12}\lambda + \Gamma^2{}_{22}\lambda^2 + \lambda'$, $r = 2d_{12}\lambda$, and simplify, we find the condition (7.7) for a hypergeodesic of (3.1) to be a plane curve is exactly the condition (7.5) for a hypergeodesic of (3.1) to be a torsal curve of (3.1). Hence

THEOREM 7.1. *Non-asymptotic hypergeodesics of a family of hypergeodesics are plane if and only if they are the torsal curves of the family. Asymptotic hypergeodesics of a family are always the torsal curves of the famliy and are plane if and only if they are straight lines.*

Equation (7.5) is in general of the sixth degree in λ and consequently has six solutions. Since λ is not zero or infinity, the vanishing of t_0 or t_6 or the vanishing of t_0, t_6 or t_0, t_1 or t_5, t_6 will reduce the degree of equation (7.5) by one or two respectively. Thus when $A = -\frac{1}{3}\Gamma^2{}_{11}$ or $D = \frac{1}{3}\Gamma^1{}_{22}$, it is in general of degree five; when $A = -\frac{1}{3}\Gamma^2{}_{11}$ and $D = \frac{1}{3}\Gamma^1{}_{22}$ or $A = -\Gamma^2{}_{11}$ or $D = \Gamma^1{}_{22}$, it is in general of degree four; when $A = -\frac{1}{3}\Gamma^2{}_{11}$ and $D = \Gamma^1{}_{22}$ or $A = -\Gamma^2{}_{11}$ and $D = \frac{1}{3}\Gamma^1{}_{22}$, it is in general of degree three; and when $A = -\Gamma^2{}_{11}$ and $D = \Gamma^1{}_{22}$, it is in general of degree two. Besides these cases in which we assume that B, C are arbitrary functions, there are other families of hypergeodesics for which B, C are also suitably chosen such that their corresponding equations (7.5) are of degree three, two, one or zero.

THEOREM 7.2. *A two-parameter family of hypergeodesics contains in general six one-parameter families of non-linear plane curves. There are* *families which contain in general five, four, three, two or one one-parameter families of such plane curves or do not contain any such plane curve. In any case, if a family contains more than the maximum number of one-parameter families of non-linear plane curve which the family is to contain in general, then all hypergeodesics of the family are plane. A family of union curves is the only family of hypergeodesics which contains in general a net of non-linear plane curves and for which B, C are arbitrary.*

A comparison among the coefficients in equations (7. 5) corresponding to two reciprocal families will yield the first part of the following theorem. The second part is seen from the fact that these equations with B, C arbitrary are both of degree two when and only when $\Gamma^2_{11} = 0$, $\Gamma^1_{22} = 0$.

THEOREM 7. 3. *The numbers of one-parameter families of non-linear plane curves contained in two reciprocal families of hypergeodesics are not related. In general they are both equal to six or are unequal. There are reciprocal families such that these numbers are both equal to five, four, three, two, one or zero. When B, C are arbitrary, two reciprocal families both contain a net of non-linear plane curves if and only if the surface is a quadric.*

I am grateful to Professor P. Sperry for her interest and suggestions during the preparation of this paper.

UNIVERSITY OF CALIFORNIA, BERKELEY.

BIBLIOGRAPHY.

[1] L. P. Eisenhart, *An introduction to differential geometry with the use of the tensor calculus*, Princeton University Press (1947).

[2] G. M. Green, "Memoir on the general theory of surfaces and rectilinear congruences," *Transactions of the American Mathematical Society*, vol. 20 (1919), pp. 79-153.

[3] E. P. Lane, *Projective differential geometry of curves and surfaces*, The University of Chicago Press (1932).

[4] P. Sperry, "Properties of a certain projectively defined two-parameter family of curves on a general surface," *American Journal of Mathematics*, vol. 40 (1918), pp. 213-224.

[5] C. E. Springer, "Dual geodesics on a surface," *Bulletin of the American Mathematics Society*, vol. 48 (1942), pp. 901-906.

ON RIEMANN METRICS OF CONSTANT CURVATURE.*

By AUREL WINTNER.

1. Let D be an open, simply-connected domain in the space of n real variables $x_1, \cdots, x_n$, and let (g_{ik}) be an n-rowed, real, symmetric matrix of functions $g_{ik} = g_{ik}(x_1, \cdots, x_n)$ which have a non-vanishing determinant and are of class C' (i. e., such as to possess continuous partial derivations of *first* order). Then the Christoffel symbols of second kind, $\Gamma_{ik}{}^j = \Gamma_{ik}{}^j(x_1, \cdots, x_n)$, exist and are continuous functions on D. Hence Levi-Civita's parallel transport, being determined by a linear system of ordinary differential equations with a continuous coefficient matrix, exists along every smooth path in D (here and in the sequel, a path $x_i = x_i(t)$ $(i = 1, \cdots, n)$ in D is called smooth if it has a continuously turning tangent).

The Riemannian geometry determined by (g_{ik}) on D will be called Euclidean if no vector is changed by parallel transportation along any closed smooth path. It will be proved that these metrics $ds^2 = g_{\alpha\beta}dk_\alpha dx_\beta$ are characterized by the following property:

If D is small enough, D possesses a one-to-one continuous mapping $y_i = y_i(x_1, \cdots, x_n)$ in which the n functions y_i are of class C' on D, have a non-vanishing Jacobian, and render every h_{ik} a constant, where $h_{ik} = h_{ik}(y_1, \cdots, y_n)$ is defined by placing $ds^2 = h_{\alpha\beta}dy_\alpha dy_\beta$.

The point in this theorem is that, in contrast to the classical theory, the machinery of a curvature tensor is now not made available by the geometrically artificial assumption of continuous *second* derivatives for the functions $g_{ik}(x_1, \cdots, x_n)$. In fact, since the latter are just of class C', the functions $\Gamma_{ik}{}^j(x_1, \cdots, x_n)$ can be nowhere differentiable. Correspondingly, the result seems to be new even in the case, $n = 2$, of the differential geometry of surfaces (to be embedded, rather than being given as embedded, into an (x, y, z)-space under the additional restriction that the binary form ds^2 is positive definite on D).

It will be shown at the end of the paper that the above theorem can be generalized so as to apply to those ds^2 which, when the $g_{ik}(x_1, \cdots, x_n)$ are of class C'', become the spaces of constant curvature. Since the $g_{ik}(x_1, \cdots, x_n)$ are supposed to be only of class C' (and of non vanishing determinant), the

---

* Received November 13, 1950.

constancy of the curvature cannot, of course, be defined in explicit terms. The substitute to be used in the C'-case is that supplied by the requirement of homogeneity, in the sense of Riemann-Helmholtz, of the metric defined by the function $g_{ik}(x_1, \cdots, x_n)$ of class C'.

In the particular case in which the x-space is Euclidean in the above-defined sense, only "the components of the affine connection," $\Gamma_{jk}{}^i = \Gamma_{kj}{}^i$, rather than the "metric tensor" g_{ik} as well, will enter into the considerations (cf. (5) below). Correspondingly, the first of the results announced above holds without the assumption of a ds^2 also. This result will be formulated and proved in Section 5.

2. Levi-Civita's system of n ordinary, linear differential equations, say $v'_i = f_{ia}(t) v_a$, is

$$v'_i = -\Gamma_{\alpha\beta}{}^i x'_\alpha v_\beta, \tag{1}$$

where $v_i = v_i(t)$ is the vector transported along a given smooth path $x_i = x_i(t)$ (of smooth parametrization), the prime denotes differentiation with respect to t, and all indices range from 1 to n. If the definition of a Euclidean (g_{ik}), as specified above, is compared with (1), it is readily seen that a given realization, $g_{\alpha\beta} dx_\alpha dx_\beta$, of a ds^2 is Euclidean on D if and only if, corresponding to every set of n numbers $v_i{}^0$ and to every point $(x_1{}^0, \cdots, x_n{}^0)$ of D, the system of n^2 linear "partial" differential equations

$$\partial v_i/\partial x_k = -\Gamma_{ka}{}^i(x_1, \cdots, x_n) v_a \tag{2}$$

(which actually is a "total" system) has exactly and/or at least one solution $v_i = v_i(x_1, \cdots, x_n)$ which is continuous, and therefore of class C', and satisfies the n initial conditions

$$v_i(x_1{}^0, \cdots, x_n{}^0) = v_i{}^0. \tag{3}$$

The standard answer to this existence question is expressed by saying that the identical vanishing of Riemann's curvature tensor (of four indices) is necessary and sufficient for the solvability of (2) under an arbitrary initial condition (3). This answer results as follows (cf., e. g., [3], pp. 22-23): A calculation shows that the curvature tensor vanishes identically if and only if

$$\partial\Gamma_{kl}{}^i/\partial x_j - \partial\Gamma_{kl}{}^i/\partial x_j = 0. \tag{4}$$

On the other hand, since the initial condition (3) is arbitrary, (4) is identical with the integrability condition, $\partial^2 v_i/\partial x_k \partial x_j = \partial^2 v_i/\partial x_j \partial x_k$, of the total system

(2). Consequently, it is sufficient to apply to (2) the general existence theorem of a total system (cf., e. g., [3], pp. 66-68) which, incidentally, is linear in the present case. But all of this depends on the assumption that the g_{ik}'s are of class C'', since otherwise the Γ's are not of class C', and therefore (4) is meaningless.

In the general case to be considered, the g_{ik}'s are of class C', hence the Γ's just continuous. What is then applicable to (2) is a recent improvement of the classical existence theorem of total systems; cf. [2]. In fact, if Theorem (II) in [2], pp. 760-761, is applied to the present particular case, (2), then the second of the two conditions (10), (11) on p. 761 in [2] reduces to $0 = 0$, while the first becomes

$$(5) \qquad \int_J (\Gamma_{jk}{}^i dx_k + \Gamma_{jl}{}^i dx_l) = \int\int_S (\Gamma_{al}{}^i \Gamma_{jk}{}^a - \Gamma_{ak}{}^i \Gamma_{jl}{}^a)\, dx_k dx_l.$$

Here the indices i, j and k, l are subject to the limitation $k \neq l$ but are otherwise arbitrary; J to the left denotes *any* small, rectifiable, positively oriented Jordan curve in that (x_k, x_l)-plane which results if those of the $n - 2$ coordinates $x_1, \cdots, x_n$ which are distinct from x_k and x_l are arbitrarily fixed within the n-dimensional domain D, and S to the right of (5) is the two-dimensional region surrounded by J; finally, the (Greek) repeated index, a, in the double integral is, but the (Roman) repeated indices, k and l, in the line integral are not, summation indices.

Accordingly, the problem is reduced to determination of those sets of continuous Christoffel functions $\Gamma_{jk}{}^i(x_1, \cdots, x_n)$ which satisfy (5) identically (on D).

It should be noted that no reference is now allowed to the pseudo-tensor character of the $\Gamma_{jk}{}^i$. In fact, the transformation rule of the $\Gamma_{jk}{}^i$ contains the *second* derivatives of the transforming function

$$(6) \qquad y_i = y_i(x_1, \cdots, x_n), \qquad (i = 1, \cdots, n).$$

3. For a single unknown function, say $y = y(x_1, \cdots, x_n)$, consider on D the system of differential equations

$$(7) \qquad \partial^2 y / \partial x_i \partial x_k = \Gamma_{ki}{}^a (x_1, \cdots, x_n) \partial y / \partial x_a,$$

where i and k range from 1 to n. Clearly, (7) is equivalent to the linear "total" system which consists of

$$(8) \qquad \partial w_k / \partial x_i = \Gamma_{ki}{}^a (x_1, \cdots, x_n) w^a$$

and

$$\partial y/\partial x_i = w_i \tag{9}$$

together (with $y_1, w_1, \cdots, w_n$ as $n+1$ unknown functions of the n independent variables $x_1, \cdots, x_n$). It will be noted that (8) is the adjoint of the system (2).

Since $\Gamma_{ik}^{a} = \Gamma_{ki}^{a}$, it is readily seen that the integrability conditions, which belong to the system (8)-(9) in the same way as (5) belongs to (2) alone, are represented just by (5). Hence, if Theorem (II) of [2] is applied to (8)-(9), it follows that (5) is necessary and sufficient for the following situation:

Corresponding to any set of $n+1$ numbers a; $c_1, \cdots, c_n$, the system (7) has at least one and/or exactly one solution $y = y(x_1, \cdots, x_n)$, of class C', satisfying the initial conditions

$$y(x_1^0, \cdots, x_n^0) = a, \qquad \partial y(x_1^0, \cdots, x_n^0)/\partial x_i = c_i, \tag{10}$$

where $(x_1^0, \cdots, x_n^0)$ is an arbitrary point of D. Actually, every such

$$y(x_1, \cdots, x_n) \text{ is of class } C''. \tag{11}$$

In fact, if $y(x_1, \cdots, x_n)$ is of class C' and satisfies (7), then (11) follows from continuity of the coefficient functions, $\Gamma_{kj}^{i} (= \Gamma_{jk}^{i})$, of (7).

4. Let (e_{ik}) denote the unit matrix and, under the assumption (5), let $y_k(x_1, \cdots, x_n)$, where $k = 1, \cdots, n$, denote that solution, $y = y(x_1, \cdots, x_n)$, of (7) for which the initial conditions (10) are as follows: $c_i = e_{ik}$, $a = a_k$ (say $a_k = 0$). Consider the transformation (6) determined by these n solutions of (7). The Jacobian of (6), being 1 at $(x_1^0, \cdots, x_n^0)$, cannot vanish near $(x_1^0, \cdots, x_n^0)$ (it could be concluded, by uniqueness, to be distinct from 0 throughout D, but this is immaterial here). Choose D about $(x_1^0, \cdots, x_n^0)$ so small that the inverse of the mapping (6) of D is unique. Finally, define on the y-image of D the functions h_{ik} by the assignment

$$g_{\alpha\beta}(x_1, \cdots, x_n)\, dx_\alpha dx_\beta \equiv h_{\alpha\beta}(y_1, \cdots, y_n)\, dy_\alpha dy_\beta. \tag{12}$$

Then the matrix (h_{ik}) is constant on the y-image of D.

In view of the remarks made before (6), the last statement is quite unexpected. It goes beyond what has been announced in Section 1, where the existence of the normal form $(h_{ik}) =$ const. of (g_{ik}) was claimed only by virtue of some transformation (6) of class C'. What is now claimed is that such a transformation can be chosen of class C''. In fact, (11) shows that the transformation (6), defined above, is of class C''. This agrees with the circumstance that, g_{ik} being a tensor, one degree of differentiability is lost by the occurrence of first derivatives in its transformation rule.

The verification of the statement, according to which $(h_{ik}) = \text{const.}$ holds by virtue of (12) and the present (6), can be omitted, precisely because the present (6) is of class C''. In fact, due to the latter circumstance, a well-known verification of $(h_{ik}) = \text{const.}$ (cf. [1], pp. 577-578) applies without any change, even though (g_{ik}) is just of class C'.

Another, more straightforward, verification of $(h_{ik}) = \text{const.}$ is contained in the more general approach to be taken in Section 5.

5. Since the integrability conditions (5), on which everything is based, contain only the Christoffel symbols, rather than the components of the metric tensor as well, it is natural to disregrad the assumption the coefficients of (1) can be derived from a symmetric $g_{ik}(x_1, \cdots, x_n)$, of class C' and of non-vanishing determinant, by placing $\Gamma_{jk}{}^i = g^{ia}\Gamma_{ajk}$, $(g^{ik}) = (g_{ik})^{-1}$ and

$$2\Gamma_{ijk} = \partial g_{ij}/\partial x_k + \partial g_{ik}/\partial x_j - \partial g_{jk}/\partial x_i.$$

All that should be given is a pseudo-tensor $\Gamma_{jk}{}^i = \Gamma_{kj}{}^i$ (cf. [3], p. 12), which is assumed to be a continuous function of the position $(x_1, \cdots, x_n)$ on D. Then (1) still defines parallel transportation, and (5) is still necessary and sufficient in order that all such transportations be independent of the path.

Under the latter assumption, and without assuming the differentiability of the Γ's, it can be concluded that there exist local transformations (6), of class C'' and of non-vanishing Jacobian, which transform the functions $\Gamma_{jk}{}^i$ of $(x_1, \cdots, x_n)$ so as to make them identically 0 in $(y_1, \cdots, y_n)$. Then (2), (1) appear in the respective normal forms $\partial v_i/\partial y_k \equiv 0$, $v_i = \text{const.}$, and so the result of Section 4 for the Riemannian case follows as a corollary.

First, since (5) is assumed, Section 3 is applicable without change. Hence, it is possible to choose the n functions (6) in the same way as in Section 4. But each of these n functions satisfies (11) and (7). It follows therefore from (9) and from the transformation rule of the Γ's that the y-transform of (2) is $\partial v_i/\partial y_k \equiv 0$.

In what follows, it will again be assumed that the geometry, instead of having just an "affine connection" (Weyl), is Riemannian.

6. If λ is a real, non-vanishing constant, let (8_λ) and (7_λ) respectively denote the systems which result if the term $-\lambda g_{ik}(x_1, \cdots, x_n)y$ is added to the right of (8) and (7). Thus (7_λ) is identical with (8_λ) by virtue of (9). As in the case $\lambda = 0$, treated above, the functions $g_{ik} = g_{ki}$ on D are supposed to be just of class C' and of non-vanishing determinant; so that the functions $\Gamma_{jk}{}^{i} = \Gamma_{kj}{}^{i}$ exist and are continuous on D.

In view of Weyl's approach to the "Helmholtz space problem" (cf. [3], pp. 24-29; also pp. 71-74), it is natural to declare the C'-metric of $g_{\alpha\beta}dx_\alpha dy_\beta$ a metric of constant curvature if, corresponding to arbitrary initial conditions (10), (3), the system consisting of (9) and (8_λ) (and belonging to a suitable $\lambda = \text{const.}$) has a solution of class C'; in other words, if the integrability conditions of (9) and (8_λ) are satisfied.

In the C''-case, this system of integrability conditions is that representation of the curvature tensor (of four indices) which is required by formula (17) in [3], p. 26, i. e., by formula (16) in [1], p. 574. In the present C'-case, where a curvature tensor cannot be written down, the integrability conditions of (8_λ)-(9) are those integral identities, extending the above (5) from $\lambda = 0$ to any $\lambda = \text{const.}$, which result if the necessary and sufficient conditions of Theorem (II), in [2], pp. 760-761, are applied to the above system (8_λ)-(9). The resulting system of integral conditions, which differ from (5) only in additive terms containing the g_{ik}'s with the constant factor λ, will be referred to as (5_λ).

7. The following theorem can now be concluded: If $g_{\alpha\beta}(x_1, \cdots, x_n)dx_\alpha dx_\beta$ and $h_{\alpha\beta}(y_1, \cdots, y_n)dy_\alpha dy_\beta$ are two real, symmetric, n-ary forms, with coefficient functions of class C' and of non-vanishing determinants, on small x- and y-domains, respectively, and if both metrics are of constant curvature (in the sense of Section 6) and belong to the same numerical value of the scalar curvature (that is, if the number λ of Section 6 is common), then the two metrics are equivalent.

By this is meant that there exists a local transformation, (6), which is of class C', of non-vanishing Jacobian, and such as to satisfy (12). Actually, it turns out that such a transformation (6) can always be chosen to be of class C''. For the Euclidean case ($\lambda = 0$), this was proved in Section 4. Let therefore $\lambda \neq 0$. Then the assertion can be restated as follows: If the $g_{ik}(x_1, \cdots, x_n)$ satisfy the above conditions, then, locally, there exist n functions (6), of class C'' and of non-vanishing Jacobian, which have the property that, by virtue of (6) and (12),

$$h_{ik}(y_1, \cdots, y_n) = (\lambda/y_1)^{-2}\epsilon_i e_{ik},$$

where (e_{ik}) is the unit matrix and $\epsilon_i = \pm 1$. The proof of this C''-assertion proceeds as follows:

According to Section 6, the constancy of the curvature of $g_{\alpha\beta}dx_\alpha dx_\beta$ means that (5_λ), being the integrability condition of the system (8_λ)-(9), is satisfied. But the latter system is identical with (7_λ). Hence, (7_λ) has a solution $y(x_1, \cdot\cdot\cdot, x_n)$, of class C', satisfying any given initial condition (10). Choose n linearly independent solutions, (6), for (7_λ) (as for (7) in Section 4). For the same obvious reason as in Section 3, each of these n solutions of (7_λ) will satisfy (11). Consequently, the classical C''-proof (cf. [1], pp. 579-581) of the assertion of the last formula line can be repeated without any change.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] L. Bianchi, *Vorlesungen über Differentialgeometrie*, Leipzig, 1899.

[2] P. Hartman and A. Wintner, "On the fundamental equations of differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 757-774.

[3] H. Weyl, *Mathematische Analyse des Raumproblems*, Berlin, 1923.

ON INDUCED REPRESENTATIONS OF GROUPS.*

By George W. Mackey.[1]

Introduction. Let $\mathfrak{G}$ be a finite group, let G be a subgroup of $\mathfrak{G}$, and let $L(\xi \rightarrow L_\xi)$ be a representation of G by non-singular linear transformations in a vector space $\mathfrak{H}(L)$ which is finite-dimensional over a field $\mathfrak{F}$. Then the functions f from $\mathfrak{G}$ to $\mathfrak{H}(L)$ such that

$$f(\xi x) = L_\xi f(x) \tag{1}$$

for all $x \in \mathfrak{G}$ and all $\xi \in G$ form a vector space over $\mathfrak{F}$ which is invariant under translation from the right by members of $\mathfrak{G}$. In other words if we define U_s^L by the equation $(U_s^L(f))(x) = f(xs)$ then U^L $(s \rightarrow U_s^L)$ will be a representation of $\mathfrak{G}$ by non-singular linear transformations in the vector space $\mathfrak{H}(U^L)$ of functions f satisfying (1). The representation U^L is called the (imprimitive) representation of $\mathfrak{G}$ "induced" by L. Induced representations have been studied by many writers beginning with Frobenius, who invented them [3]. It is the purpose of part I of this article to unify a part of the work of Frobenius and his successors by showing that a number of their results are easy consequences of a single theorem (Theorem 2) about Kronecker products of induced representations. Specifically, Theorem 2 implies the Frobenius reciprocity theorem [3], Shoda's results [10] on the irreducibility and equivalence of monomial representations and Artin's theorem [1] to the effect that any character is a rational linear combination of characters of induced representations of a very special sort.

Some of our results are extensible to Hilbert space representations of locally compact topological groups. In part these extensions involve rather complicated topological and measure theoretic considerations. Certain of them however involve a minimum of such analysis and these are presented in part II. Our more involved results will appear in another paper addressed primarily to analysts.

---

* Received October 10, 1950.

[1] Most of the work on which this paper is based was done while the author was a fellow of the John Simon Guggenheim Memorial Foundation in residence successively at the University of Chicago, the University of Nancy, and the Institute for Advanced Study.

I. Finite Groups.

1. Definitions and preliminary lemmas. We shall adhere to the notation introduced in the introduction and in particular shall always use $\mathcal{H}(U)$ to denote the vector space of a representation U. If $\mathcal{H}$ is any vector space over $\mathcal{F}$ we shall denote by $\bar{\mathcal{H}}$ the conjugate space; that is the space of all linear functions from $\mathcal{H}$ to $\mathcal{F}$. If T is a linear transformation of $\mathcal{H}$ into itself, T^* will denote the linear transformation of $\bar{\mathcal{H}}$ into itself such that $T^*(l)(v) = l(T(v))$ for all l in $\bar{\mathcal{H}}$ and all v in $\mathcal{H}$.

If $U(x \rightarrow U_x)$ is any representation of $\mathfrak{G}$ we shall denote by $\bar{U}$ and call the *adjoint* of U the representation $x \rightarrow U_{x^{-1}}{}^*$. Clearly $\mathcal{H}(\bar{U}) = \overline{\mathcal{H}(U)}$.

Let $U(x \rightarrow U_x)$ and $V(y \rightarrow V_y)$ be representations of the groups $\mathfrak{G}_1$ and $\mathfrak{G}_2$. Let $\mathcal{H}_1$ be the set of all linear transformations from $\overline{\mathcal{H}(V)}$ to $\mathcal{H}(U)$. For each $x, y \,\varepsilon\, \mathfrak{G}_1 \times \mathfrak{G}_2$, $T \rightarrow U_x T V_y{}^*$ is a linear transformation of $\mathcal{H}_1$ into itself. If we denote this by $U_x \times V_y$ then $x, y \rightarrow U_x \times V_y$ is a representation of $\mathfrak{G}_1 \times \mathfrak{G}_2$ in $\mathcal{H}_1$. We shall call this representation of $\mathfrak{G}_1 \times \mathfrak{G}_2$ the *outer Kronecker product* $U \times V$ of U and V. Now let U and V be representations of the same group $\mathfrak{G}$. $\mathfrak{G} \times \mathfrak{G}$ then contains a subgroup $\tilde{\mathfrak{G}}$ isomorphic to $\mathfrak{G}$; namely the group of all x, y with $x = y$. $U \times V$ restricted to $\tilde{\mathfrak{G}}$ then defines a representation of $\mathfrak{G}$. We call this the *Kronecker product* $U \otimes V$ of U and V.

Let L be a representation of the subgroup G of $\mathfrak{G}$ and consider $\mathcal{H}(U^L)$. Let l' be any member of $\overline{\mathcal{H}(U^L)}$. Then as is easy to see there exists a function l'' $(x \rightarrow l_x'')$ from $\mathfrak{G}$ to $\overline{\mathcal{H}(L)}$ such that $l'(f) = \Sigma_{x \varepsilon G} l_x''(f(x))$ for all $f \,\varepsilon\, \mathcal{H}(U^L)$. But $\Sigma_{x \varepsilon G} l_x''(f(x)) = \Sigma_{G/G} \Sigma_{\xi \varepsilon G} l_{\xi x}''(f(\xi x))$, where the second sum is over the right-G-cosets. Now

$$\Sigma_{\xi \varepsilon G} l_{\xi x}''(f(\xi x)) = \Sigma_{\xi \varepsilon G} L_\xi^*(l_{\xi x}'')(f(x)).$$

Thus if we set $l_x = \Sigma_{\xi \varepsilon G} L_\xi^*(l_{\xi x}'')$ we have $l_{\xi x} = L_{\xi^{-1}}{}^*(l_x)$ for all $\xi \,\varepsilon\, G$ and all $x \,\varepsilon\, \mathfrak{G}$ and

$$l(f) = \Sigma_{G/G} l_x(f(x)). \tag{2}$$

The reader will now have no difficulty in verifying the truth of

LEMMA 1. *The mapping of $\mathcal{H}(U^{\bar{L}})$ into $\overline{\mathcal{H}(U^L)}$ defined by equation (2) is an isomorphism onto and sets up an equivalence between the representations $U^{\bar{L}}$ and $\overline{U^L}$.*

Next let G_1 and G_2 be subgroups $\mathfrak{G}_1$ and $\mathfrak{G}_2$ respectively and let L and M be representations of G_1 and G_2. Let A be any member of $\mathcal{H}(U^{L \times M})$. Then A is a function from $\mathfrak{G}_1 \times \mathfrak{G}_2$ to $\mathcal{H}(L \times M)$ such that

$A_{\xi x, \eta y} = (L_\xi \times M_\eta)(A_{x,y})$ for all $x, y \in \mathfrak{G}_1 \times \mathfrak{G}_2$ and all $\xi, \eta \in G_1 \times G_2$. In other words $A_{x,y}$ is for each x, y a linear transformation from $\overline{\mathcal{H}(M)}$ to $\mathcal{H}(L)$ and $A_{\xi x, \eta y} = L_\xi A_{x,y} M_\eta{}^*$ for all $\xi, \eta \in G_1 \times G_2$ and all $x, y \in \mathfrak{G}_1 \times \mathfrak{G}_2$. Thus if l $(y \to l_y)$ is any member of $\mathcal{H}(U^{\bar{M}})$ we see that $A_{x,y}(l_y)$ as a function of y depends only upon the right G_2-coset to which y belongs and that we may define a member f of $\mathcal{H}(U^L)$ as follows:

$$f(x) = \Sigma^{y}_{\mathfrak{G}_2/G_2} A_{x,y}(l_y). \tag{3}$$

By Lemma 1 there is a natural map of $\mathcal{H}(U^{\bar{M}})$ on $\overline{\mathcal{H}(U^M)}$. Thus (3) may be regarded as defining a linear map of $\overline{\mathcal{H}(U^M)}$ on $\mathcal{H}(U^L)$; that is, a member of $\mathcal{H}(U^L \times U^M)$. Let us denote this member of $\mathcal{H}(U^L \times U^M)$ by $A^{\sim}$. The reader can now readily satisfy himself of the truth of

LEMMA 2. *The mapping $A \to A^{\sim}$ of $\mathcal{H}(U^{M \times L})$* into *$\mathcal{H}(U^L \times U^M)$ is an isomorphism onto and sets up an equivalence between the representations $U^L \times U^M$ and $U^{L \times M}$.*

2. The main theorems. Let now G_1 and G_2 be two subgroups of the group $\mathfrak{G}$ and let L be a representation of G_1. Let us denote by ${}_{G_2}U^L$ the restriction to G_2 of the representation U^L of $\mathfrak{G}$. We observe at once that ${}_{G_2}U^L$ admits a direct sum decomposition into as many parts as there are double cosets $G_1 x G_2$. Indeed let $x_1, x_2, \cdots, x_n$ be a choice of elements, one from each double coset, so that $\mathfrak{G} = \bigcup_{i=1}^{n} G_1 x_i G_2$. For each $i = 1, 2, \cdots, n$ and each $f \in \mathcal{H}(U^L)$ let $f_i(x) = f(x)$ if $x \in G_1 x_i G_2$ and let $f_i(x) = 0$ if $x \notin G_1 x_i G_2$. Then f_i is also in $\mathcal{H}(U^L)$ and the set of all f_i for $f \in \mathcal{H}(U^L)$ is precisely the set $\mathcal{H}_i$ of all $f \in \mathcal{H}(U^L)$ which vanish outside of $G_1 x_i G_2$. Finally it is clear that for each $s \in G_2$, $U_s{}^L$ carries each $\mathcal{H}_i$ into itself. Thus $\mathcal{H}(U^L) = \mathcal{H}_1 \oplus (\mathcal{H}_2 \oplus \cdots \oplus \mathcal{H}_n$ defines a direct sum decomposition of ${}_{G_2}U^L$. Let us examine the representation associated with $\mathcal{H}_i$. If f is any member of $\mathcal{H}_i$ we may define a function $\tilde{f}$ from G_2 to $\mathcal{H}(L)$ as follows: $\tilde{f}(t) = f(x_i t)$. Since $f(\xi x_i t) = L_\xi f(x_i t) = L_\xi \tilde{f}(t)$ for all $\xi \in G_1$ and $t \in G_2$, it is clear that $\tilde{f}$ determines f completely and that $f \to \tilde{f}$ is a one-to-one map. Let $x_i^{-1} \xi x_i = \eta$ be any element of $G_2 \cap x_i^{-1} G_1 x_i$. Then $\tilde{f}(\eta t) = f(\xi x_i t) = L_\xi f(x_i t) = L_{x_i \eta x_i^{-1}}(\tilde{f}(t))$. Conversely, if g is any function from G_2 to $\mathcal{H}(L)$ such that $g(\eta t) = L_{x_i \eta x_i^{-1}}(g(t))$ for all $\eta \in G_2 \cap x_i^{-1} G_1 x_i$ and all $t \in G_2$, then it may be verified at once that $f(\xi x_i t) = L_\xi(g(t))$ defines f unambiguously as a member of $\mathcal{H}_i$ such that $g = \tilde{f}$. In other words $f \to \tilde{f}$ sets up an equivalence between the component of ${}_{G_2}U^L$ in $\mathcal{H}_i$ and the representation of G_2 induced by the representation $\eta \to L_{x_i \eta x_i^{-1}}$ of the subgroup $G_2 \cap x_i^{-1} G_1 x_i$. Thus we have proved

THEOREM 1. *Let* L, G_1, G_2 *and* $\mathfrak{G}$ *be described above. For each* $x \varepsilon \mathfrak{G}$ *consider the subgroup* $G_2 \cap x^{-1}G_1x$ *of* G_2 *and let* V^x *denote the representation of* G_2 *induced by the representation* $\eta \rightarrow L_{x\eta x^{-1}}$ *of this subgroup. Then* V^x *is determined to within equivalence by the double coset* $G_1xG_2 = D(x)$ *to which* x *belongs and we may write* $V^D = V^x$ *where* $D = D(x)$. U^L *restricted to* G_2 *is the direct sum of the* V^D *over the double* $G_1 : G_2$ *cosets* D.

As an almost immediate consequence of Theorem 1 and Lemma 2 we have

THEOREM 2.[2] *Let* G_1 *and* G_2 *be subgroups of* $\mathfrak{G}$ *and let* L *and* M *be representations of* G_1 *and* G_2 *respectively. For each* $x, y \varepsilon \mathfrak{G} \times \mathfrak{G}$ *consider the representations* $s \rightarrow L_{xsx^{-1}}$ *and* $s \rightarrow M_{ysy^{-1}}$ *of the subgroup* $(x^{-1}G_1x) \cap (y^{-1}G_2y)$ *of* $\mathfrak{G}$. *Let us denote their Kronecker product by* $N^{x,y}$ *and form the induced representation* $U^{N^{x,y}}$ *of* $\mathfrak{G}$. *Then* $U^{N^{x,y}}$ *is determined to within equivalence by the double coset* $G_1xy^{-1}G_2$ *to which* xy^{-1} *belongs and the direct sum of the* $U^{N^{x,y}}$ *over the double cosets is equivalent to the Kronecker product* $U^L \otimes U^M$ *of* U^L *and* U^M.

Proof. $U^L \otimes U^M$ is the representation of $\mathfrak{G}$ obtained from the representation $U^L \times U^M$ *of* $\mathfrak{G} \times \mathfrak{G}$ by restriction to the isomorphic replica $\tilde{\mathfrak{G}}$ of $\mathfrak{G}$ consisting of all $x, y \varepsilon \mathfrak{G} \times \mathfrak{G}$ with $x = y$. Moreover by Lemma 2, $U^L \times U^M$ is equivalent to $U^{L \times M}$, where $L \times M$ is of course a representation of $G_1 \times G_2$. By Theorem 1, $U^{L \times M}$ restricted to $\tilde{\mathfrak{G}}$ is a direct sum over the double cosets $(G_1 \times G_2)(x, y)\tilde{\mathfrak{G}}$ and the summand associated with the double coset containing x, y is the representation of $\tilde{\mathfrak{G}}$ induced by the representation $s, s \rightarrow (L \times M)_{(x,y)(s,s)(x,y)^{-1}}$ of $\tilde{\mathfrak{G}} \cap (x, y)^{-1}(G_1 \times G_2)(x, y)$. But it is easily verified that $\tilde{\mathfrak{G}} \cap (x, y)^{-1}(G_1 \times G_2)(x, y)$ transferred to $\mathfrak{G}$ by the natural isomorphism is the subgroup $x^{-1}G_1x \cap y^{-1}G_2y$ and the the representation $s, s \rightarrow (L \times M)_{(x,y)(s,s)(x,y)^{-1}}$ becomes $s \rightarrow L_{xsx^{-1}} \otimes M_{ysy^{-1}}$. Finally a simple calculation shows that $x_1, y_1 \varepsilon (G_1 \times G_2)(x, y)\tilde{\mathfrak{G}}$ if and only if $x_1y_1^{-1} \varepsilon G_1xy^{-1}G_2$.

Given two representations U and V of $\mathfrak{G}$ let us define the intertwining number $\mathfrak{J}(U, V)$ of U and V to be the dimension of the vector space of all linear transformations T from $\mathcal{H}(V)$ to $\mathcal{H}(U)$ such that $U_xT = TV_x$ for all $x \varepsilon \mathfrak{G}$. As is evident and well known, $\mathfrak{J}(U, V)$ is equal to the number of times that the Kronecker product $U \otimes V$ contains the identity; that is to the dimension of the set of all $w \in \mathfrak{H}(U \otimes V)$ such that $(U \otimes V)_x(w) = w$ for all $x \in \mathfrak{G}$. Moreover for any induced representation U^L the number of times that U^L contains the identity is equal to the number of times that L contains the identity. Indeed, if $f \in \mathfrak{H}(U^L)$ then $U^L(f) = f$ if and only if f is a constant function on $\mathfrak{G}$, and if $f(x) = w$ for all $x \in \mathfrak{G}$ then $f \in \mathfrak{H}(U^L)$ if and only if $L_\xi(w) = w$ for all $\xi \in G_1$. From these remarks and Theorem 2 we deduce at once

---

[2] We are indebted to the referee for pointing out that the special case of Theorem 2 in which $G_2 = \mathfrak{G}$ and L is the identity representation of G_1 has been obtained by M. Osima [9]. The referee reports further that Prof. Osima later obtained more general results—one he believes quite close to ours—but that so far as he knows these results have not been published.

THEOREM 3. *Let $\mathfrak{G}$, G_1, G_2, L and M be as in Theorem 2. For each $x, y \in \mathfrak{G}$ consider the representations $s \rightarrow L_{xsx^{-1}}$ and $s \rightarrow M_{ysy^{-1}}$ of $x^{-1}G_1x \cap y^{-1}G_2y$, and let $\mathfrak{J}(L, M, x, y)$ denote their intertwining number. Then $\mathfrak{J}(L, M, x, y)$ depends only upon the double coset $D = D(x, y) = G_1xy^{-1}G_2$ to which xy^{-1} belongs, so that we may write $\mathfrak{J}(L, M, D)$. Finally $\mathfrak{J}(U^L, U^M) = \Sigma_{D \in \mathfrak{D}} \mathfrak{J}(L, M, D)$, where $\mathfrak{D}$ is the class of $G_1 : G_2$ double cosets.*

3. Applications. Our applications of Theorem 3 depend upon the following well-known property of intertwining numbers. Let U and V be representations of $\mathfrak{G}$ such that U is irreducible and let $\mathfrak{F}$ be algebraically closed and such that its characteristic does not divide the order of $\mathfrak{G}$. Then V is a direct sum of irreducible parts and $\mathfrak{J}(U, V)$ is equal to the number of parts which are equivalent to U. We deduce at once

THEOREM 4. (Frobenius) *Let G_1 be a subgroup of $\mathfrak{G}$ and let L and M be irreducible representations of G_1 and $\mathfrak{G}$ respectively. Let $\mathfrak{F}$ be algebraically closed and of characteristic prime to the order of $\mathfrak{G}$. Then the number of times that U^L contains M as a direct summand is equal to the number of times that M restricted to G_1 contains L as a direct summand.*

Proof. In Theorem 3 take $G_2 = \mathfrak{G}$ and observe that there is only one double coset.

As a corollary of Theorems 3 and 4 we have

THEOREM 5. (Frobenius) *Let $\mathfrak{F}$ be as in Theorem 4 and let G_1 and G_2 be subgroups of $\mathfrak{G}$. For each $i = 1, 2$ and each irreducible representation V^γ $(\gamma = 1, 2, \cdots, \sigma)$ of $\mathfrak{G}$ let $r_\gamma{}^i$ be the number of times that V^γ restricted to G_i contains the identity representation. Then $\Sigma_{\gamma=1}^{\sigma} r_\gamma{}^1 r_\gamma{}^2$ is equal to the number of $G_1 : G_2$ double cosets.*

Proof. By Theorem 4, $\Sigma_{\gamma=1}^{\sigma} r_\gamma{}^1 r_\gamma{}^2 = \mathfrak{J}(U^L, U^M)$, where L (resp. M)

is the identity representation of G_1 (resp. G_2). By Theorem 3, $\mathfrak{I}(U^L, U^M)$ is in this case the number of $G_1 : G_2$ double cosets.

THEOREM 6. *Let $\mathfrak{G}$, G_1, and $\mathfrak{F}$ be as in Theorem 4. Then U^L is irreducible if and only if for each x not in G_1 the representations of $x^{-1}G_1x \cap G_1$ defined by $s \rightarrow L_{xsx^{-1}}$ and $s \rightarrow L_s$ have no common irreducible components.*

Proof. Observe that a representation is irreducible if and only if its intertwining number with itself is 1 and apply Theorem 3 with $G_2 = G_1$ and $L = M$.

Theorem 6 for the special case in which L is one-dimensional is due to Shoda [10]. In this case its statement is slightly simpler for the two representations $s \rightarrow L_{xsx^{-1}}$ and $s \rightarrow L_s$ are one-dimensional and hence fail to have common irreducible components if and only if they are distinct. Our next corollary of Theorem 3 is also due to Shoda [10] in the special case in which L and M are one-dimensional.

THEOREM 7. *Let $\mathfrak{G}$, G_1, G_2, L and M be as in Theorem 2 and let $\mathfrak{F}$ be as in Theorem 4. Let U^L and U^M be irreducible. Then U^L and U^M are equivalent if and only if there exists $x \in \mathfrak{G}$ such that the representations $s \rightarrow L_{xsx^{-1}}$ and $s \rightarrow M_s$ of the subgroup $x^{-1}G_1x \cap G_2$ have an irreducible component in common.*

For our next application we shall need the stronger Theorem 2. We begin with a definition and a lemma. Let G_1 be a cyclic subgroup of $\mathfrak{G}$ and let L be a one-dimensional representation of G_1. Then we shall call U^L an Artin representation of $\mathfrak{G}$ and its character $\chi(U^L, x) = \text{Trace}\,(U_x{}^L)$ an Artin character.

LEMMA. *If $\mathfrak{F}$ is as in Theorem 4 and x_1 and x_2 are in different conjugate classes in $\mathfrak{G}$ then there exists an Artin character χ such that $\chi(x_1) \neq \chi(x_2)$.*

Proof. Let G_1 be the cyclic group generated by x_1, let its order be n and let ω be a primitive n-th root of unity. Then the general one-dimensional representation of G_1 is obtained by choosing $k = 0, 1, \cdots, n-1$ and setting $L_k(x_1{}^r) = \omega^{kr}$ for $r = 1, 2, \cdots, n$. Let the classes containing x_1 and x_2 be denoted by C_1 and C_2 and let C_1, C_2 and $\mathfrak{G}$ contain h_1, h_2 and h elements respectively. From the well-known and easily verified formula $\chi(U^L, x) = (1/n)\Sigma_{s \in G}\phi(s^{-1}xs)$, where $\phi(\xi) = \chi(L, \xi)$ or zero according as ξ is or is not in G_1 and where L is a representation of G_1, we compute at once that

$$\chi(U^{L_k}, x_1) = (1/n)(h/h_1)\Sigma_{x_1^r \varepsilon C_1}\omega^{kr} = \Sigma_{x_1^r \varepsilon C_1}(h/h_1 n)\omega^{kr}$$

and

$$\chi(U^{L_k}, x_2) = (1/n)(h/h_2)\Sigma_{x_1^r \varepsilon C_2}\omega^{kr} = \Sigma_{x_1^r \varepsilon C_2}(h/h_2 n)\omega^{kr}.$$

But the mappings $\theta \rightarrow \psi$, where $\psi(k) = \Sigma_{r=0}^{n-1}\theta(r)\omega^{kr}$, and $\psi \rightarrow \theta$, where $\theta(r) = \Sigma_{k=0}^{n-1}\omega^{-kr}\psi(k)$, set up a one-to-one correspondence between $\mathfrak{F}$-valued functions θ and $\mathfrak{F}$-valued functions ψ. Thus $\chi(U^{L_k}, x_1) \neq \chi(U^{L_k}, x_2)$ for at least one k. This establishes the truth of the lemma.

As a consequence of this lemma and Theorem 2 we may now prove

THEOREM 8. (Artin) *If $\mathfrak{F}$ is as in Theorem 4 then any character χ on $\mathfrak{G}$ is a finite linear combination with rational coefficients of Artin characters.*

Proof. Let $\mathfrak{A}$ be the set of all class functions on $\mathfrak{G}$ which are finite linear combinations with coefficients in $\mathfrak{F}$ of Artin characters. It is an immediate corollary of Theorem 2 that the Kronecker product of two Artin representations is a direct sum of Artin representations and hence that the product of two Artin characters is a linear combination with positive integral coefficients of Artin characters. Thus $\mathfrak{A}$ under multiplication is an algebra over $\mathfrak{F}$. Moreover by the immediately preceding lemma it contains functions distinguishing between any two classes. It follows at once that $\mathfrak{A}$ contains all class functions with values in $\mathfrak{F}$. Thus if there are c classes in $\mathfrak{G}$ there must exist c Artin characters $\chi_1, \chi_2, \cdots, \chi_c$ which form a basis for $\mathfrak{A}$ over $\mathfrak{F}$. Let $\chi_1^0, \chi_2^0, \cdots, \chi_c^0$ be the irreducible characters of $\mathfrak{G}$. Then we have $\chi_i = \Sigma_{j=1}^{c} n_{ij}\chi_j^0$, where $\| n_{ij} \|$ is a matrix of integers. Since the χ_i form a basis, $\| n_{ij} \|$ has an inverse $\| r_{ij} \|$. Since the n_{ij} are integers, the r_{ij} are rational and the truth of the theorem is evident.

It is to be remarked that Artin proved the somewhat stronger result that the only prime factors occurring in the denominators of the r_{ij} are divisors of h and that Brauer [2] has since shown that by admitting certain non-cyclic subgroups one can dispense altogether with non-integral rational numbers. Whether the methods of this paper can be used to obtain new proofs of these results is not known.

II. Infinite Groups.

1. Preliminaries. From now on the group $\mathfrak{G}$ will be a separable locally compact topological group. By a unitary representation U $(x \rightarrow U_x)$ of $\mathfrak{G}$ we shall mean a homomorphism of $\mathfrak{G}$ into the group of all unitary transformations of some Hilbert space $\mathfrak{H}(U)$ into itself. We shall consider only unitary representations which are continuous in the sense that for each $v \varepsilon \mathfrak{H}(U)$ the mapping $x \rightarrow U_x(v)$ from $\mathfrak{G}$ to $\mathfrak{H}(U)$ is continuous and we shall use the words "unitary representation" to mean continuous unitary representation. We remark that in order to ensure the continuity of a U it is enough to ensure that for each v and w in $\mathfrak{H}(U)$ the function $U_x(v) \cdot w$ of x is a Borel function. Here the dot indicates the scalar product in the Hilbert space. Two unitary representations U and V will be said to be unitary equivalent if there exists a unitary transformation A from $\mathfrak{H}(U)$ to $\mathfrak{H}(V)$ such that $AU_xA^{-1} = V_x$ for all x in $\mathfrak{G}$.

The definition of U^L where L is a unitary representation of the closed subgroup G of $\mathfrak{G}$ is rather complicated in the general case. See [5]. Here we shall have occasion to deal only with subgroups G which are open as well as closed so that the space $\mathfrak{G}/G$ of right-G-cosets is denumerable and discrete. In this case we may define U^L as follows. Consider first the vector space of all functions f from $\mathfrak{G}$ to $\mathfrak{H}(L)$ such that $f(\xi x) = L_\xi f(x)$ for all $\xi \varepsilon G$ and $x \varepsilon \mathfrak{G}$. For each such f we have $f(\xi x) \cdot f(\xi x) = L_\xi f(x) \cdot L_\xi f(x) = f(x) \cdot f(x)$ since L_ξ is unitary. Thus $f(x) \cdot f(x)$ depends only upon the right-G-coset to which x belongs. Let us set $\| f \|^2 = \Sigma_{G/G} f(x) \cdot f(x)$, where the sum is over the right-G-cosets. We define $\mathfrak{H}(U^L)$ to be the set of all f in the space under consideration for which $\| f \|^2 < \infty$. We leave to the reader the routine task of verifying that $\mathfrak{H}(U^L)$ is indeed a Hilbert space under the norm just defined, that $f \cdot g = \Sigma_{G/G} f(x) \cdot g(x)$, and that if $U_s{}^L(f)(x) = f(xs)$ then $s \rightarrow U_s{}^L$ is a unitary representation of $\mathfrak{G}$.

Consider now the problem of carrying over the theory of part I. If we try to do this more or less directly we are confronted with a dilemma. Suppose we define the space of the Kronecker produce of U and V as the set of all bounded linear operators from $\overline{\mathfrak{H}(V)}$ to $\mathfrak{H}(U)$. Then not only do we not get a Hilbert space but we find that the obvious generalizations of Lemma 2 and Theorem 2 are false. Suppose on the other hand that we secure a Hilbert space for $\mathfrak{H}(U \otimes V)$ by restricting attention to those operators form $\overline{\mathfrak{H}(V)}$ to $\mathfrak{H}(U)$ which are in the Schmidt class. Then we may prove analogues of Theorems 1 and 2 and even of Theorem 3 if we define intertwining numbers using only operators in the Schmidt class. Unfortunately, however, an analogue of Theorem 3 involving such "strong" intertwining numbers does not have implications about irreducibility and unitary equivalence. Thus when we pass to the infinite case we lose some of the unity of part I and must consider separately Kronecker products and restrictions to subgroups on the one hand and irreducibility, unitary equivalence and ordinary intertwining numbers on the other. Here we shall treat only the second topic and that in the special case in which the subgroups under consideration are both open and closed. It turns out to be possible, in this case at least, to prove an analogue of Theorem 3 without using Kronecker products. Our results will be found summarized in [6]. Our results on the first topic will be treated in detail elsewhere. They are summarized in [7].

2. The main theorem. As suggested above we define an intertwining operator T for the representations U and V of $\mathfrak{G}$ as a bounded linear operator from $\mathfrak{H}(V)$ to $\mathfrak{H}(U)$ such that $U_xT = TV_x$ for all x in $\mathfrak{G}$. The intertwining number $\mathfrak{J}(U, V)$ is then defined as the dimension $(= 0, 1, 2, \cdots, \infty)$ of the vector space of all such intertwining operators. We wish to compute $\mathfrak{J}(U^L, U^M)$ where L and M are unitary representations of the open closed subgroups G_1 and G_2 of $\mathfrak{G}$.

Suppose for the moment that $\mathfrak{G}$ is finite and that the representations considered are finite-dimensional, so that the methods of part I may be applied. The fact that the intertwining operators of U^L and U^M are defined by members of the space of $U^L \otimes \overline{U^M}$ and that the latter is an isomorphic image of the space of $U^{L \times \bar{M}}$ tells us that every intertwining operator is defined by a function A from $\mathfrak{G} \times \mathfrak{G}$ to the space of all linear transformations from $\mathfrak{H}(M)$ to $\mathfrak{H}(L)$ such that $A_{\xi x,\eta y} = L_\xi A_{x,y} M_\eta{}^*$ for all $\xi,\eta \in G_1 \times G_2$. Conversely, every such function A defines a linear transformation from $\mathfrak{H}(U^M)$ to $\mathfrak{H}(U^L)$ as follows. Consider $A_{x,y}(f(y))$, where $f \in \mathfrak{H}(U^M)$. Then for each $\eta \in G_2$, $A_{x,\eta y}(f(\eta y)) = A_{x,y}M_\eta{}^* M_\eta(f(y)) = A_{x,y}(f(y))$. Thus for fixed x, $A_{x,y}(f(y))$ depends only upon the right-G_2-coset to which y belongs. We may sum over these right-cosets and indicate this by $\overset{y}{\Sigma}_{G/G_2} A_{x,y}(f(y))$. The resulting function of x is in $\mathfrak{H}(U^L)$. The transformation $A^{\sim}$ from $\mathfrak{H}(U^M)$ to $\mathfrak{H}(U^L)$ thus defined is an intertwining operator if and only if $A_{xs,ys} = A_{x,y}$ for all $s \in \mathfrak{G}$. In short $\mathfrak{J}(U^L, U^M)$ is the dimension of the set of all functions A such that $A_{\xi xs,\eta ys} = L_\xi A_{x,y} M_\eta{}^*$ for all $\xi \in G_1$, $\eta \in G_2$, $s \in \mathfrak{G}$. Theorem 3 follows fairly readily from this fact.

Let us return now to the infinite case. Using the openness of G_1 and G_2 it is easy to show that every intertwining operator is defined by a function A on $\mathfrak{G} \times \mathfrak{G}$ of the sort just described. Unfortunately not every such function defines an intertwining operator and it has not proved possible to characterize those which do in a useful manner. There does turn out however to exist a conveniently characterizable subclass of functions A which can be proved to have the same dimension as the class of those which do define intertwining operators. Our analogue of Theorem 3 follows from a determination of the dimension of this class. We proceed now to the details of these matters.

LEMMA A. *Let T be any intertwining operator for the unitary representations U^L and U^M of $\mathfrak{G}$; L and M being unitary representations of the open and closed subgroups G_1 and G_2 of $\mathfrak{G}$. Then there exists a function A $(x, y \rightarrow A_{x,y})$ from $\mathfrak{G} \times \mathfrak{G}$ to the class of bounded linear operators from $\mathcal{H}(M)$ to $\mathcal{H}(L)$ such that*

(a) $A_{\xi xs, \eta ys} = L_\xi A_{x,y} M_\eta{}^*$ *for all* $\xi \varepsilon G_1$, $\eta \varepsilon G_2$, $s \varepsilon \mathfrak{G}$.

(b) *there exists a constant* $K > 0$ *such that*

$$\sum^{x}_{G/G_1} \| A_{x,y}(v) \|^2 / \| v \|^2 \leq K \text{ for all } y \varepsilon \mathfrak{G} \text{ and all } v \varepsilon \mathcal{H}(M),$$

(c) *there exists a constant* $K' > 0$ *such that*

$$\sum^{y}_{G/G_2} \| A_{x,y}{}^*(v) \|^2 / \| v \|^2 \leq K' \text{ for all } x \varepsilon \mathfrak{G} \text{ and all } v \varepsilon \mathcal{H}(L),$$

(d) *for each* $f \varepsilon \mathcal{H}(U^M)$, $(T(f))(x) = \sum^{y}_{G/G_2} A_{x,y}(f(y))$,

(e) *for each* $f \varepsilon \mathcal{H}(U^L)$, $(T^*(f))(x) = \sum^{x}_{G/G_1} A_{x,y}{}^*(f(y))$,

the summation in (b), (c), (d), *and* (e) *being over the right-cosets indicated.*

Proof. Given an element y of $\mathfrak{G}$ and an element v of $\mathcal{H}(M)$, let $f_{y,v}(\eta y) = M_\eta(v)$ for all $\eta \varepsilon G_2$ and let $f_{y,v}(x) = 0$ for all x not in the right-coset $G_2 y$. Then $f_{y,v}$ is clearly a member of $\mathcal{H}(U^M)$. Moreover any member g of $\mathcal{H}(U^M)$ which is zero outside of $G_2 y$ is of the form $f_{y,v}$ with $v = g(y)$. Finally every member of $\mathcal{H}(U^M)$ is a unique sum over the right-G_2-cosets of members vanishing outside of these cosets. Thus the intertwining operator T is completely determined by its values $T(f_{y,v})$ at the $f_{y,v}$. Consider now $(T(f_{y,v}))(x)$. It is a member of $\mathcal{H}(L)$ which for fixed x and y depends linearly on v. Let us denote this linear transformation from $\mathcal{H}(M)$ to $\mathcal{H}(L)$ by $A_{x,y}$, so that $A_{x,y}(v) = (T(f_{y,v}))(x)$. Now $\| f_{y,v} \| = \| v \|$, so $\| T(f_{y,v}) \| \leq \| T \| \| v \|$. Thus $\| A_{x,y}(v) \| \leq \| T \| \| v \|$ for all x and y, and $A_{x,y}$ is seen to be a bounded linear transformation. We shall leave to the reader the straightforward computations necessary to verify equations (a), (d) and (e). Inequality (b) is an immediate consequence of the fact that T is bounded and that $T(f_{y,v})$ is in $\mathcal{H}(U^L)$. Indeed, $\| T(f_{y,v}) \|^2$

$=\Sigma_{G/G_1}\| A_{x,y}(v)\|^2 \leq \| T \|^2 \| f_{y,v} \|^2 = \| T \|^2 \| v \|^2$ and we may take $K = \| T \|^2$. Inequality (c) follows on applying the same argument to T^*.

LEMMA B. *$\mathcal{I}(U^L, U^M)$ is equal to the dimension d of the vector space of all functions A having the properties* (a), (b) *and* (c) *listed in Lemma* A.

Proof. It follows at once from Lemma A that $d \geq \mathcal{I}(U^L, U^M)$. Thus if $\mathcal{I}(U^L, U^M) = \infty$ the equality holds trivially. It will suffice then to prove that whenever A satisfies (a), (b) and (c) and $\mathcal{I}(U^L, U^M) < \infty$ then A defines an intertwining operator via equation (d). Let A have these properties. Let $\mathcal{M}$ be the subspace of $\mathcal{H}(U^M)$ consisting of all $f \, \varepsilon \, \mathcal{H}(U^M)$ which vanish outside of a finite number of right-G_2-cosets. $\mathcal{M}$ is dense in $\mathcal{H}(U^M)$. Since all but a finite number of the terms vanish there is no difficulty in forming the sum $\overset{y}{\Sigma}_{G/G_2}A_{x,y}(f(y))$ for each $f \, \varepsilon \, \mathcal{M}$. This sum as a function of x is then an element of $\mathcal{H}(U^L)$. We obtain in this way a linear operator T_0 defined in $\mathcal{M} \subseteq \mathcal{H}(U^M)$ and having values in $\mathcal{H}(U^L)$. Similarly $g \to \overset{x}{\Sigma}_{G/G_1}A_{x,y}{}^*(g(x))$ defines a linear operator T_1 whose arguments form a dense subspace $\mathcal{L}$ of $\mathcal{H}(U^L)$ and whose values are in $\mathcal{H}(U^M)$. We verify at once that T_0 and T_1 are adjoint to one another; that is, $T_0(f) \cdot g = f \cdot T_1(g)$ for all $g \, \varepsilon \, \mathcal{L}$ and $f \, \varepsilon \, \mathcal{M}$. Hence $T_0{}^{**}$ exists and is a closed linear operator extending T_0. Moreover $T_0{}^*$ is a closed linear operator extending T_1. Now it follows at once from the definition of T_0 that $U_x{}^L T_0 = T_0 U_x{}^M$ for all $x \, \varepsilon \, \mathcal{G}$. Hence

$$(*) \qquad T_0{}^* \overset{L}{U}_{x^{-1}} = \overset{M}{U}_{x^{-1}} T_0{}^*$$

$$(**) \qquad U_x{}^L T_0{}^{**} = T_0{}^{**} U_x{}^M.$$

We shall suppose for simplicity that $T_0{}^{**}$ is one-to-one and has its range dense in $\mathcal{H}(U^L)$. Whenever this is not the case we need only replace U^M by its restriction to the orthogonal complement of the null space of $T_0{}^{**}$ and U^L by its restriction to the closure of the range of $T_0{}^{**}$ and apply the following argument to these two new representations. Making this assumption then about $T_0{}^{**}$, we may write $T_0{}^{**} = WH$, where W is a unitary map of $\mathcal{H}(U^M)$ on $\mathcal{H}(U^L)$ and H is non-negative and self-adjoint. Thus $T_0{}^* = HW^*$, so that $T_0{}^* T_0{}^{**} = H^2$. Multiplying together equations (*) and (**) we find that $T_0{}^* T_0{}^{**} = \overset{M}{U}_{x^{-1}} T_0{}^* T_0{}^{**} U_x{}^M$. Thus H^2 and hence H commutes with all $U_x{}^M$. Substituting $T_0{}^{**} = WH$ in (**) and using the fact that H commutes with all $U_x{}^M$ we find that $(U_x{}^L W - W U_x{}^M) H = 0$ and hence that $U_x{}^L W = W U_x{}^M$ for all x. It follows at once that if $K = (1 + H)^{-1}$

then WK^n is an intertwining operator for U^L and U^M for all $n = 0, 1, 2, \cdots$. Since, by hypothesis, $\mathfrak{J}(U^L, U^M)$ is finite, only a finite number of the WK^n can be linearly independent. It follows that the spectrum of K is finite and hence that H must be bounded. Thus T_0^{**} and hence T_0 is bounded. Finally then T_c has a unique bounded extension to all of $\mathfrak{H}(U^M)$ and this extension is an intertwining operator. Lemma B is thus established.

Lemma C. *Let* $\tilde{\mathfrak{G}}$ *be the subgroup of* $\mathfrak{G} \times \mathfrak{G}$ *consisting of all* x, y *for which* $x = y$. *For each double coset* $D = (G_1 \times G_2)(x, y)\tilde{\mathfrak{G}}$ *let* d_D *be the dimension of the set of all functions* A *which satisfy condition* (a), (b) *and* (c) *of Lemma* A *and which vanish outside of* D. *Then* $d = \Sigma_{D \varepsilon \mathcal{D}} d_D$, *where* $\mathcal{D}$ *is the family of all* $(G_1 \times G_2) : \mathfrak{G}$ *double cosets.*

Proof. It is obvious that a function A which satisfies (a) is uniquely determined throughout $(G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$ by its value A_{x_0, y_0} at x_0, y_0 and it is also obvious that an A which satisfies (b) and (c) as well will continue to do so after having been reduced to zero outside of this double coset.

Lemma D. *Let* d'_D *be the dimension of the set of all functions* A *which satisfy condition* (a) *of Lemma* A *and which vanish outside of the double coset* D. *Let* x, y *be any point of* D. *Then* d'_D *is equal to the intertwining number of the representations* $s \rightarrow L_{xsx^{-1}}$ *of the subgroup intertwining number of the representations* $s \rightarrow L_{xsx^{-1}}$ *and* $s \rightarrow M_{ysy^{-1}}$ *of the subgroup* $x^{-1}G_1x \cap y^{-1}G_2y$ *of* $\mathfrak{G}$.

Proof. A is completely determined by its value at x, y. If this value is B then $A_{\xi xs, \eta ys}$ will be equal to $L_\xi B M_\eta^*$ for all $\xi \varepsilon G_1$, $\eta \varepsilon G_2$ and $s \varepsilon \mathfrak{G}$. Conversely, given a B it will be $A_{x,y}$ for some A if and only if the expression $L_\xi B M_\eta^*$ depends only upon ξxs and ηys and not upon ξ and η; that is, if and only if $\xi_1 x s_1 = \xi_2 x s_2$ and $\eta_1 y s_1 = \eta_2 y s_2$ imply $L_{\xi_1} B M_{\eta_1}^* = L_{\xi_2} B M_{\eta_2}^*$; that is, if and only if $\xi_2^{-1}\xi_1 x = x s_2 s_1^{-1}$ and $\eta_2^{-1}\eta_1 y = y s_2 s_1^{-1}$ imply $L_{\xi_2^{-1}\xi_1} B = B M_{\eta_2^{-1}\eta_1}$; that is, if and only if $x^{-1}\xi x = s$ and $y^{-1}\eta y = s$ imply $L_\xi B = B M_\eta$; that is, if and only if $s \varepsilon x^{-1}G_1x \cap y^{-1}G_2y$ implies $L_{xsx^{-1}} B = B M_{ysy^{-1}}$; that is, if and only if B is an intertwining operator for the representations of the lemma.

Lemma E. *Let* L *and* M *be one-dimensional and let* A *be a function satisfying* (a), (b) *and* (c) *of Lemma* A. *Then if* $A_{x_0, y_0} \neq 0$ *the index of* $x_0^{-1}G_1x_0 \cap y_0^{-1}G_2y_0$ *is finite in both of the containing subgroups* $x_0^{-1}G_1x_0$ *and* $y_0^{-1}G_2y_0$. *Conversely, if these two indices are finite then every* A *which satisfies* (a) *of Lemma* A *and vanishes outside of* $(G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$ *also satisfies* (b) *and* (c). *This last assertion holds without restriction on the dimensions of* L *and* M.

Proof. If L and M are one-dimensional then A_{x_0,y_0} is a number which is constant in absolute value throughout the double coset to which x_0, y_0 belongs. Now $\| A_{x_0,y_0}(v) \|^2 / \| v \|^2$ is simply the square of this number. Thus condition (b) can be fulfilled only if there are only finitely many summands; that is, only finitely many right-G_1-cosets $G_1 x$ for which $x, y_0 \varepsilon (G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$. Similarly condition (c) can be satisfied only if there are only finitely many right-G_2-cosets $G_2 y$ for which $x_0, y \varepsilon (G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$. But $x, y_0 \varepsilon (G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$ if and only if there exists $\xi \varepsilon G_1$, $\eta \varepsilon G_2$ and $s \varepsilon \mathfrak{G}$ such that $x = \xi x_0 s$ and $y_0 = \eta y_0 s$; that is, if and only if $\xi \varepsilon G_1$ and $\eta \varepsilon G_2$ exist so that $x^{-1} \xi x_0 = y_0^{-1} \eta y_0$, or $x = \xi x_0 y_0^{-1} \eta^{-1} y_0$; that is, if and only if $x \varepsilon (G_1 x_0)(y_0^{-1} G_2 y_0)$. Thus the number of right-G_1-cosets $G_1 x$ for which $x, y \varepsilon (G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$ is equal to the number of right-G_1-cosets in $(G_1 x_0)(y_0^{-1} G_2 y_0)$. But $\xi_1 x_0 y_0^{-1} \eta_1 y_0$ and $\xi_2 x_0 y_0^{-1} \eta_2 y_0$ belong to the same G_1-right-coset if and only if ξ exists in G_1 with $\xi \xi_1 x_0 y_0^{-1} \eta_1 y_0 = \xi_2 x_0 y_0^{-1} \eta_2 y_0$; that is, $x_0^{-1} \xi_2^{-1} \xi \xi_1 x_0 = y_0^{-1} \eta_2 \eta_1^{-1} y_0$; that is, $y_0^{-1} \eta_2 \eta_1^{-1} y_0 \varepsilon x_0^{-1} G_1 x_0$. Thus the number of right-G_1-cosets in $(G_1 x_0)(y_0^{-1} G_2 y_0)$ is equal to the index of $y_0^{-1} G_2 y_0 \cap x_0^{-1} G_1 x_0$ in $y_0^{-1} G_2 y_0$, and if condition (b) is satisfied this index must be finite. Similarly if condition (c) is satisfied the same argument shows that the index of this group in $x_0^{-1} G_1 x_0$ must be finite. Conversely, if these two indices are finite then for arbitrary L and M there are only a finite number of summands in (b) and (c). Since these numbers are independent of v and y the desired inequalities are evidently satisfied. This completes the proof of the lemma.

As the reader may easily prove for himself the two indices under discussion in the preceding lemma depend only upon the double coset $(G_1 \times G_2)(x_0, y_0)\tilde{\mathfrak{G}}$ to which x_0, y_0 belongs. Moreover x_0, y_0 and x_1, y_1 belong to the same $(G_1 \times G_2) : \tilde{\mathfrak{G}}$ double coset if and only if $x_0 y_0^{-1}$ and $x_1 y_1^{-1}$ belong to the same $G_1 : G_2$ double coset. Combining these observations with Lemmas A, B, C, D and E we obtain at once

THEOREM 3′. *Let G_1 and G_2 be both open and closed subgroups of the separable locally compact group $\mathfrak{G}$. Let L and M be unitary representations of G_1 and G_2 respectively. For each x and y in $\mathfrak{G}$ consider the representations $s \rightarrow L_{xsx^{-1}}$, $s \rightarrow M_{ysy^{-1}}$ of $x^{-1} G_1 x \cap y^{-1} G_2 y$ and let $\mathfrak{J}(L, M, x, y)$ denote their intertwining number. Then $\mathfrak{J}(M, L, x, y)$ depends only upon the double coset $D = D(x, y) = G_1 x y^{-1} G_2$ to which xy^{-1} belongs, so that we may write $\mathfrak{J}(L, M, D)$. Moreover the indices of $x^{-1} G_1 x \cap y^{-1} G_2 y$ in $x^{-1} G_1 x$ and in $y^{-1} G_2 y$ depend only upon this double coset. Let $\mathfrak{D}_f$ be the set of all double* *cosets for which both indices are finite and let* $\mathfrak{D}$ *be the set of all double cosets. Then*

$$\Sigma_{D\varepsilon\mathfrak{D}_f}\mathcal{J}(L,M,D) \leq \mathcal{J}(U^L,U^M) \leq \Sigma_{D\varepsilon\mathfrak{D}}\,\mathcal{J}(L,M,D);$$

and if L *and* M *are one-dimensional then*

$$\Sigma_{D\varepsilon\mathfrak{D}_f}\mathcal{J}(L,M,D) = \mathcal{J}(U^L,U^M).$$

3. Applications. Theorems 4′, 6′ and 7′ are deducible from Theorem 3′ just as were Theorems 4, 6, and 7 from Theorem 3. We leave details to the reader.

THEOREM 4′. *Let* G_1 *be a closed and open subgroup of the separable locally compact group* $\mathfrak{G}$. *Let* L *and* M *be irreducible unitary representations of* G_1 *and* $\mathfrak{G}$ *respectively. Then the number of times that* U^L *contains* M *as a discrete direct summand is less than or equal to the number of times that the restriction of* M *to* G_1 *contains* L *as a discrete direct summand. If* $\mathfrak{G}/G_1$ *is finite then this inequality may be replaced by an equality.*

THEOREM 6′. *Let* $\mathfrak{G}$, G_1 *and* L *be as in Theorem 4′ and let* L *be one-dimensional. Then* U^L *is irreducible if and only if for each* x *not in* G_1 *one of the following statements is true.*

(a) L_s *and* $L_{xsx^{-1}}$ *are not identical for* s *in* $x^{-1}G_1x \cap G_1$.

(b) $x^{-1}G_1x \cap G_1$ *has infinite index in at least one of the two intersecting subgroups.*

We remark here that condition (b) is slightly misstated in [6].

THEOREM 7′ *Let* $\mathfrak{G}$, G_1, G_2, L *and* M *be as in Theorem 3′. Let* L *and* M *be one-dimensional and let* U^L *and* U^M *be irreducible. Then* U^L *and* U^M *are unitary equivalent if and only if there exists* $x \,\varepsilon\, \mathfrak{G}$ *such that*

(a) M_s *and* $L_{xsx^{-1}}$ *are identical for* s *in* $x^{-1}G_1x \cap G_2$ *and*

(b) $x^{-1}G_1x \cap G_2$ *has finite index in each intersecting subgroup.*

When L is the identity representation of G_1 then condition (a) of Theorem 6′ is never satisfied. Since condition (b) is independent of L we deduce at once

THEOREM 9. *Let* $\mathfrak{G}$ *and* G_1 *be as in Theorem 6′. A necessary and sufficient condition that* U^L *be irreducible for all one-dimensional representations* L *of* G_1 *is that condition* (b) *of Theorem 6′ hold.*

Theorem 9 in the slightly more special case for which $\mathfrak{G}$ is discrete and G_1 Abelian is due to Godement [4].

We conclude this section with a brief discussion of a class of examples. Let G_1 and G_2 be infinite discrete Abelian groups. Let there be given a homomorphism of G_2 into the group of automorphisms of G_1 and let $x \rightarrow y[x]$ denote the automorphism of G_1 defined by $y \varepsilon G_2$. Let $\mathfrak{G}$ be the set of all pairs x, y with $x \varepsilon G_1$, $y \varepsilon G_2$ and let us convert $\mathfrak{G}$ into a group by defining multiplication as follows: $(x_1, y_1)(x_2, y_2) = (x_1 y_1[x_2], y_1 y_2)$. Then the set of all x, e where e is the identity of G_2 is a group isomorphic with G_1 and may be identified with G_1. Similarly the set of all e, y may be identified with G_2. $\mathfrak{G}$ is a "semi-direct product" of G_1 and G_2.

LEMMA 1. *Let G_1 and G_2 be such that all non-trivial orbits of G_1 under G_2 are infinite. Then for each one-dimensional representation M of G_2 the representation U^M is irreducible. Moreover U^M and U^L are unitary equivalent if and only if M and L are the same one-dimensional representation of G_2.*

Proof. A straightforward computation shows that $(x, e)^{-1} G_2 (x, e)$ is the set of all $(x^{-1} y[x], y)$ for $y \varepsilon G_2$. Thus $(x, e)^{-1} G_2 (x, e) \cap G_2$ is the set of all $y \varepsilon G_2$ with $y[x] = x$. Its index in G_2 is thus the number of distinct transforms of x by elements of y. By hypothesis, then, this index is infinite for $x \neq e$. Since x, e and x, y are in the same $G_2 : G_2$ double coset we conclude that condition (b) of Theorem 6′ holds. The lemma is now seen to be a consequence of Theorems 6′ and 7′.

LEMMA 2. *If L and M are one-dimensional representations of G_1 and G_2 respectively then $\mathfrak{J}(U^L, U^M) = 0$.*

Proof. Since $G_1 G_2 = \mathfrak{G}$ there is only one double coset and it is clearly not in $\mathfrak{D}_f$.

LEMMA 3. *If L is a one-dimensional representation of G_1 then U^L is irreducible if and only if $L(y[x]) \neq L(x)$ for all y in G_2 distinct from the identity.*

Proof. The truth of the lemma is an immediate consequence of Theorem 6′.

It follows from a theorem in [4] (stated more generally in [5]) that the regular representation of $\mathfrak{G}$ is a direct integral over the character group $\hat{G}_1$ of G_1 of the representations U^L for $L \varepsilon \hat{G}_1$ and also is a direct integral over the character group $\hat{G}_2$ of G_2 of the representations U^M for $M \varepsilon \hat{G}_2$.

Let us now choose as G_1 the additive group of rational numbers and as G_2 the multiplicative group of non-zero rational numbers and let $y[x]$ be the product of x and y. Applying Lemmas 1, 2, and 3 we may conclude that for almost all $L \varepsilon \hat{G}_1$, U^L is irreducible, that for all $M \varepsilon \hat{G}_2$, U^M is irreducible, and that no U^M is unitary equivalent to any U^L or any other U^M. Thus we have

THEOREM 10. *There exists an infinite discrete group whose regular representation may be decomposed into infinite-dimensional irreducible parts in such a manner that no two are unitary equivalent.*

THEOREM 11. *There exists an infinite discrete group whose regular representation may be decomposed into infinite-dimensional irreducible parts in two*[3] *entirely different ways; specifically, so that no component in one decomposition is unitary equivalent to any component in the other.*

4. Concluding remarks. Theorem 3′ is evidently quite far from being a best possible result. The following questions immediately present themselves. (1) Can the inequality be replaced by an equality when L and M are not one-dimensional? (2) Can a result like Theorem 3′ be proved when G_1 and G_2 are not assumed to be open? (3) Can a version of Theorem 3′ be given which uses a less crude definition of intertwining number—one which takes cognizance of the fact that a representation may decompose as a direct integral of irreducible representations rather than as a discrete direct sum of such. Questions (2) and (3) are now under investigation and the results will be reported upon at a later time. Question (1) we have considered only to the extent of observing that the final statement of Theorem 3′ is certainly false unless some restriction is put on L and M. Indeed if in the example just considered we let L and M be the regular representations of G_1 and G_2 we find that $\mathfrak{D}_f$ is empty while $\mathfrak{I}(U^L, U^M) = \infty$.

HARVARD UNIVERSITY.

Added in proof. An independent example of the phenomenon described in Theorem 11 has been found by H. Yoshizawa, "Some remarks on unitary representations of the free group," *Osaka Mathematical Journal,* vol. 3 (1951), pp. 56-63.

---

[3] For a special class of semi-direct products one of these two decompsitions has been considered by F. I. Mautner [8].

BIBLIOGRAPHY.

[1] Artin, E., "Über eine neue Art von L-Reihen," *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, vol. 3 (1924), pp. 89-108.

[2] Brauer, R., "On Artin's L-series with general group characters," *Annals of Mathematics*, vol. 48 (1947), pp. 502-514.

[3] Frobenius, G., "Über Relationen zwischen den Characteren einer Gruppe und denen ihrer Untergruppen," *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 1898, pp. 501-515.

[4] Godement, R., "Sur la transformation de Fourier dans les groupes discrets," *Comptes Rendus de l'Académie des Sciences* (*Paris*), vol. 228 (1949), pp. 627-628.

[5] Mackey, G. W., "Imprimitivity for representations of locally compact groups I," *Proceedings of the National Academy of Sciences, U. S. A.*, vol. 35 (1949), pp. 537-545.

[6] ———, "Imprimitivité pour les représentations des groupes localement compacts. II. Nombres d'entrelacement pour les représentations imprimitives," *Comptes Rendus de l'Académie des Sciences* (*Paris*), vol. 230 (1950), pp. 808-809.

[7] ———, "Imprimivité pour les représentations des groupes localement compacts. III. Produits de Kronecker et nombres d'entralacement forts," *Comptes Rendus de l'Académie des Sciences* (*Paris*), vol. 230 (1950), pp. 908-909.

[8] Mautner, F. I., "Completeness relations on locally compact groups," *Bulletin of the American Mathematical Society*, vol. 55 (1949), p. 52.

[9] Osima, M., "Note on the Kronecker product of representaitons of a group," *Proceedings of the Imperial Academy of Tokyo*, vol. 17 (1941), pp. 411-413.

[10] Shoda, K., "Über die monomialen Darstellungen einer endlichen Gruppe," *Proceedings of the Physico-Mathematical Society of Japan*, vol. 15 (1933), pp. 249-257.

THE RECIPROCITY FORMULA FOR DEDEKIND SUMS.*

By L. J. Mordell.

Let p, q be two positive integers without a common divisor. It is well known that

$$\sum_{x=1}^{p-1}[qx/p]+\sum_{y=1}^{q-1}[py/q]=(p-1)(q-1), \tag{1}$$

where $[z]$ denotes the integer part of z. There exists another result of this type discovered by Dedekind in discussing the linear transformation of the modular function $\log\eta(\omega)$, one form of which is

$$q\sum_{x=1}^{p-1}x[qx/p]+p\sum_{y=1}^{q-1}y[py/q]=\tfrac{1}{12}(p-1)(q-1)(8pq-p-q-1). \tag{2}$$

Rademacher [2] made a detailed study of this result and has published some five proofs. One is a joint proof with Whiteman, and the last has only just appeared. Some of them are arithmetical in character and quite simple. Another proof has just been given by Rédei [3], and a generalization by Apostol [1]. I notice, however, an entirely different way of considering the subject which is no less simple and relates the result to more general and obvious ones.

Let us consider the sum

$$S=\sum_{K}(qx+py) \tag{3}$$

extended over the integer sets (x, y) or say the lattice points P lying in the region K defined by

$$0<x<p,\quad 0\leq y<q,\quad qx+py<pq,$$

and so if O, A, B are the points $(0,0)$, $(p,0)$, $(0,q)$ respectively, K is the open triangle OAB. We call K' the open triangle ACB where C is the point (p,q). We have a 1-1 correspondence between the lattice points $P(x,y)$ in K and $P'(x',y')$ in K' given by

$$x+x'=p,\qquad y+y'=q.$$

Since K and K' together contain $(p-1)(q-1)$ lattice points, K contains exactly $\frac{1}{2}(p-1)(q-1)$ lattice points.

---

* Received December 5, 1950.

It is well known and obvious that the formula (1) states that the number of lattice points in $K + K'$ is the sum of those in the open triangles OAC and OBC.

In (3), we note that

$$\sum_K x = \sum_{x=1}^{p-1} x[q - qx/p] = \sum_{x=1}^{p-1} (p - x)[qx/p]$$

since $[q - qx/p]$ values of y correspond to given x. Hence

$$\begin{aligned} S &= q \sum_{x=1}^{p-1} (p - x)[qx/p] + p \sum_{y=1}^{q-1} (q - y)[py/q] \\ &= pq\Big(\sum_{x=1}^{p-1} [qx/p] + \sum_{y=1}^{q-1} [py/q]\Big) - \sum_{x=1}^{p-1} qx[qx/p] - \sum_{y=1}^{q-1} py[py/q]. \end{aligned}$$

Then from (1) and (2), we have

$$S = pq(p-1)(q-1) - \tfrac{2}{3}pq(p-1)(q-1) + \tfrac{1}{12}(p-1)(q-1)(p+q+1),$$

and so

$$S = \tfrac{1}{3}pq(p-1)(q-1) + \tfrac{1}{12}(p-1)(q-1)(p+q+1). \tag{4}$$

Thus the proof of (2) is reduced to the evaluation of the sum (3) as given in (4).

We solve now the more general problem of evaluating

$$T = \sum_K f(qx + py) \tag{5}$$

where f is an arbitrary polynomial and the summation is extended over the lattice points in K.

Write $\xi = qx + py$ for lattice points x, y in K so that ξ is not divisible by p or q and $0 < \xi < pq$. Then the numbers $pq - \xi$ cannot be represented in this way. For

$$\text{if } pq - \xi = qX + pY, \text{ then } pq = q(x + X) + p(y + Y),$$

and so $x + X \equiv 0 \pmod{p}$. Then $x + X = p$ and similarly $y + Y = q$, and this is clearly impossible. The number of ξ is $\tfrac{1}{2}(p-1)(q-1)$ and so the ξ and $pq - \xi$ are $(p-1)(q-1)$ in number, and so together they are precisely the integers X not divisible by p or q in the interval $0 < X < pq$.

There is, however, such a representation for the numbers $2pq - \xi$. Write $\xi' = qx' + py'$ for lattice points x', y' in K' so that $pq < \xi' < 2pq$. We have now a 1-1 correspondence between the representation of ξ and ξ' given by

$$x' + x = p, \qquad y' + y = q, \qquad \xi' + \xi = 2pq.$$

We prove two fundamental formulae. The first is

$$(7)\qquad \sum_{y=1}^{q-1}\sum_{x=1}^{p-1} f(qx+py) = \sum_K f(\xi) + \sum_K f(2pq-\xi).$$

The left hand side consists of two parts corresponding to $qx + py < pq$, and to $pq < qx + py < 2pq$. The first is $\sum_K f(\xi)$, and on putting $x = p - x'$, $y = q - y'$, the second is $\sum_K f(2pq - \xi)$.

We prove next that

$$(8)\qquad \sum_{X=1}^{pq-1} f(X) - \sum_{X=1}^{p-1} f(qX) - \sum_{X=1}^{q-1} f(pX) = \sum_K f(\xi) + \sum_K f(pq-\xi).$$

The left hand side of this is $\sum f(Y)$ extended over the numbers Y in $0 < Y < pq - 1$ and not divisible by p or q. These numbers Y can be written as ξ or $pq - \xi$ since the numbers Y in $0 < Y < pq - 1$ which cannot be represented by ξ are given by $pq - \xi$. This gives (8).

The two equations (7), (8) determine $\sum_K f(\xi)$ for any polynomial $f(\xi)$. Take $f(\xi) = \xi^2$, then (7) becomes

$$\sum_{y=1}^{q-1}\sum_{x=1}^{p-1} (qx+py)^2 = \sum_K \xi^2 + \sum_K (2pq-\xi)^2$$

and so

$$\begin{aligned}\tfrac{1}{6}q^2(q-1)p(p-1)(2p-1) &+ \tfrac{1}{6}p^2(p-1)q(q-1)(2q-1)\\ &+ \tfrac{1}{2}p^2q^2(p-1)(q-1)\\ &= 2\sum_K \xi^2 - 4pq\sum_K \xi + 2p^2q^2(p-1)(q-1),\end{aligned}$$

since K contains $\tfrac{1}{2}(p-1)(q-1)$ lattice points.

Hence

$$\begin{aligned}(9)\qquad \tfrac{1}{6}q(q-1)(p-1)(2p-1) &+ \tfrac{1}{6}p(p-1)(q-1)(2q-1)\\ &- \tfrac{3}{2}pq(p-1)(q-1) = 2/(pq)\sum_K \xi^2 - 4\sum_K \xi.\end{aligned}$$

Next (8) becomes

$$\sum_{X=1}^{pq-1} X^2 - q^2\sum_{X=1}^{p-1} X^2 - p^2\sum_{X=1}^{q-1} X^2 = \sum_K \xi^2 + \sum_K (pq-\xi)^2$$

or

$$\begin{aligned}\tfrac{1}{6}pq(pq-1)(2pq-1) &- \tfrac{1}{6}q^2p(p-1)(2p-1)\\ &- \tfrac{1}{6}p^2q(q-1)(2q-1)\\ &= 2\sum_K \xi^2 - 2pq\sum_K \xi + \tfrac{1}{2}p^2q^2(p-1)(q-1).\end{aligned}$$

This becomes

$$(10)\quad \tfrac{1}{6}(pq-1)(2pq-1)-\tfrac{1}{6}q(p-1)(2p-1)-\tfrac{1}{6}p(q-1)(2q-1)$$
$$-\tfrac{1}{2}pq(p-1)(q-1)$$
$$=2/(pq)\sum_K \xi^2-2\sum_K \xi.$$

These two equations (9), (10) determine $\sum_K \xi$, $\sum_K \xi^2$ and give the result (4). Thus (9) is

$$(11)\quad \tfrac{1}{6}(p-1)(q-1)(-5pq-p-q)=2/(pq)\sum_K \xi^2-4\sum_K \xi,$$

and (10) is

$$(12)\quad \tfrac{1}{6}(p-1)(q-1)(-pq+1)=2/(pq)\sum_K \xi^2-2\sum_K \xi.$$

In fact the left hand side of (10) vanishes when $p=1$ or $q=1$, so we can write it as $\tfrac{1}{6}(p-1)(q-1)(apq+b(p+q)+c)$, where a, b, c are constants.

Equating terms in p^2q^2, $p+q$, 1, clearly

$$\frac{a}{6}=\frac{2}{6}-\frac{1}{2},\ \frac{c}{6}=\frac{1}{6},\ \frac{b}{6}-\frac{c}{6}=-\frac{1}{6},$$

and so $a=-1$, $c=1$, $b=0$.

Hence from (11), (12)

$$(13)\quad 2\sum_K \xi=\tfrac{1}{6}(p-1)(q-1)(4pq+p+q+1)$$

$$(14)\quad 2/(pq)\sum_K \xi^2=\tfrac{1}{6}(p-1)(q-1)(3pq+p+q+2).$$

The result (13) is the required result (4).

It is also clear that on taking $f(\xi)=\xi^{2n}$ in (7), (8), the equations determine $\sum_K \xi^{2n}$, $\sum_K \xi^{2n-1}$ when we know the values of $\sum_K \xi^r$, $0\leqq r\leqq 2n-2$.

If, however, in (7) we replace the function $f(\xi)$ by $f(\xi-pq)$ and subtract from (8), we have

$$(15)\quad \sum_K f(\xi)-\sum_K f(\xi-pq)$$
$$=\sum_{X=1}^{pq-1} f(X)-\sum_{X=1}^{q-1} f(pX)-\sum_{X=1}^{p-1} f(qX)-\sum_{y=1}^{q-1}\ \sum_{x=1}^{p-1} f(qx+py-pq).$$

Write in the usual notation for the Bernouillian polynomial $B_n(x)$,

$$te^{tx}/(e^t-1)=\sum_{n=0}^{\infty}B_n(x)t^n/n!, \tag{16}$$

so that

$$te^{tx}=\sum_{n=0}^{\infty}(B_n(x+1)-B_n(x))t^n/n!, \qquad B_n(x+1)-B_n(x)=nx^{n-1}.$$

If we take $f(X)=B_n(1+X/(pq))$ in (15), we have the explicit formula for $\sum_K \xi^{n-1}$. This takes the shape

$$n\sum_K \xi^{n-1}=B_n(pq)-p^{n-1}B_n(q)-q^{n-1}B_n(p)+(pq)^{n-1}B_n(1)$$
$$-\sum_{y=1}^{q-1}\sum_{x=1}^{p-1}B_n(x/p+y/q),$$

on noting that

$$\sum_{r=0}^{p-1}B_n(x+r/p)=B_n(px)/p^{n-1}.$$

(as remarked to me by Professor Rademacher).

We can also find an explicit formula for

$$p^{n+1}q^{n+1}\sum_K B_n(x/p+y/q)$$

as a polynomial in p and q. One of a different type has been given by Apostol for odd n by using in a different way lattice points in a triangle.

As well known, $B_n(X)=(-1)^nB_n(1-X)$, from (16). Hence when n is even, we have at once from (8),

$$2\sum_K B_n(x/p+y/q)=\sum_{X=1}^{pq-1}B_n(X/pq)-\sum_{X=1}^{p-1}B_n(X/p)-\sum_{X=1}^{q-1}B_n(X/q).$$

To find the result for all n, write (15) as

$$\sum_K f(\xi/pq)-\sum_K f(\xi/pq-1) \tag{17}$$
$$=\sum_{X=1}^{pq-1}f(X/pq)-\sum_{X=1}^{p-1}f(X/p)-\sum_{X=1}^{q-1}f(X/q)-\sum_{Y=1}^{q-1}\sum_{X=1}^{p-1}f(X/p+Y/q-1).$$

Write $B_n(X)=b_nX^n+b_{n-1}X^{n-1}+\cdots+b_0$. Take

$$f(X-1)=(b_n/\{n+1\})B_{n+1}(X)+\cdots+(b_0/1)B_1(X).$$

Then

$$f(\xi/pq)-f(\xi/pq-1)=b_n/\{n+1\}(B_{n+1}(\xi/pq+1)-B_{n+1}(\xi/pq))$$
$$+\cdots=b_n(\xi/pq)^n+b_{n-1}(\xi/pq)^{n-1}+\cdots=B_n(\xi/pq).$$

Since we can put $\xi = qx + py$, the result is given by (17) on summing for X etc.

UNIVERSITY OF PENNSYLVANIA.

REFERENCES.

[1] T. M. Apostol, "Generalized Dedekind sums and transformation formulae of certain Lambert series," *Duke Mathematical Journal*, vol. 17 (1950), pp. 147-157.

[2] H. Rademacher, "Zur Theorie der Modulfunktionen," *Journal für die reine und angewandte Mathematik*, vol. 167 (1931), pp. 312-336.

———, "Eine arithmetische Summenformel," *Monatshefte für Mathematik und Physik*, vol. 39 (1932), pp. 221-228.

———, "Egy reciprocitásképletről a modulfüggvények elméletéből," (with German résumé), *Matematikai és Fizikai Lapok*, vol. 40 (1933), pp. 24-34.

———, "Die Reziprozitätsformel für Dedekindsche Summen," *Acta Scientiarum Mathematicarum*, vol. 12, part B (1950), pp. 57-60.

——— and A. Whitman, "Theorems on Dedekind sums," *American Journal of Mathematics*, vol. 63 (1941), pp. 377-407.

[3] L. Rédei, "Elementarer Beweis und Verallgemeinerung einer Reziprozitätsformel von Dedekind," *Acta Scientiarum Mathematicarum*, vol. 12, part B (1950), pp. 236-239.

THE EVALUATION OF THE CONSTANT IN THE FORMULA FOR THE NUMBER OF PARTITIONS OF *n*.*

By D. J. NEWMAN.

In 1918 Hardy and Ramanujan,[1] using the powerful circle method, proved that the number $p(n)$ of partitions of n satisfies

$$p(n) \sim a n^{-1} \exp(\pi (2n/3)^{\frac{1}{2}}), \tag{1}$$

where

$$a = 1/(4.3^{\frac{1}{2}}). \tag{2}$$

In 1942 Erdös[2] proved (1), but not (2), by completely elementary methods. The purpose of this paper is to provide an elementary proof of (2), assuming the truth of (1); this together with Erdös' proof gives the full Hardy-Ramanujan theorem without recourse to complex variable methods.

We shall begin by proving

$$\sum_{n=1}^{\infty} n^{-1} \exp(\pi (2n/3)^{\frac{1}{2}} x^n \sim 2(6/\pi)^{\frac{1}{2}} \phi(x) \tag{3}$$

and

$$\sum_{n=1}^{\infty} p(n) x^n \sim (2\pi)^{-\frac{1}{2}} \phi(x), \tag{4}$$

where

$$\phi(x) = (1-x)^{\frac{1}{2}} \exp\left(-\frac{\pi^2}{12} + \frac{\pi^2}{6(1-x)}\right),$$

and the asymptotic relations are meant to hold as x tends to 1 from below. From (1), (3), (4),

$$(2\pi)^{-\frac{1}{2}} \phi(x) \sim \sum p(n) x^n \sim \sum a n^{-1} \exp(\pi (2n/3)^{\frac{1}{2}}) x^n \sim 2(6/\pi)^{\frac{1}{2}} a \phi(x).$$

A comparison of the first and last of these four asymptotically equivalent functions leads at once to (2).

If we write $x = e^{-u}$ and stipulate that in all subsequent asymptotic

---

* Received March 8, 1950; revised November 29, 1950.

[1] G. H. Hardy and S. Ramanujan, "Asymptotic formulae in combinatory analysis," *Proceedings of the London Mathematical Society*, ser. 2, vol. 17 (1918), pp. 75-115.

[2] P. Erdös, "On an elementary proof of some asymptotic formulas in the theory of partitions," *Annals of Mathematics*, vol. 43 (1942), pp. 437-450.

formulae x tends to unity from below (*i. e.*, u tends to zero from above), then since $u = \log(1/x) \sim 1 - x$, we have $(1-x)^{\frac{1}{2}} \sim u^{\frac{1}{2}}$. Moreover,

$$\frac{1}{1-x} - \frac{1}{2} \sim \Bigl(\sum_{n=1}^{\infty} n^{-1}(1-x)^{n}\Bigr)^{-1} = \left(\log \frac{1}{1-(1-x)}\right)^{-1} = w^{-1},$$

so that $\phi(x) \sim u^{\frac{1}{2}} \exp(\pi^2/6u) = \Phi(u)$, say. Hence (3) and (4) may be written

$$\sum_{n=1}^{\infty} n^{-1} \exp(\pi(2n/3)^{\frac{1}{2}} - nu) \sim 2(6/\pi)^{\frac{1}{2}}\Phi(u), \tag{5}$$

$$\sum_{n=1}^{\infty} p(n) e^{-nu} \sim (2\pi)^{-\frac{1}{2}}\Phi(u), \tag{6}$$

and it remains only to prove (5) and (6).

We will make use of the following simple result.

LEMMA.[3] *If $f(t)$ is of bounded variation on the positive real axis, then for any $u > 0$,*

$$\Bigl| u \sum_{n=1}^{\infty} f(nu) - \int_0^{\infty} f(t)\,dt \Bigr| \leqq uV,$$

where V is the total variation of f.

Proof of (5). The sum on the left side of (5) is clearly asymptotic to

$$\begin{aligned} &2 \sum_{n=1}^{\infty} n^{-1} e^{-nu} [\sin h(\pi(2n/3)^{\frac{1}{2}}) - \pi(2n/3)^{\frac{1}{2}}] \\ &\quad = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} 2n^{-1} e^{-nu} (2\pi^2 n/3)^{m+\frac{1}{2}}/(2m+1)\,! \\ &\quad = 2 \sum_{m=1}^{\infty} (u^{-m-\frac{1}{2}}(2\pi^2/3)^{m+\frac{1}{2}}/(2m+1)\,!)\,(u \sum_{n=1}^{\infty} (un)^{m-\frac{1}{2}} e^{-nu}) \\ &\quad = 2 \sum_{m=1}^{\infty} (u^{-m-\frac{1}{2}}(2\pi^2/3)^{m+\frac{1}{2}}/(2m+1)\,!)\Gamma(m+\tfrac{1}{2})\,(1+O(u)), \end{aligned}$$

by the lemma cited above and the fact that the total variation of $t^{m-\frac{1}{2}}e^{-t}$ is

$$2(m-\tfrac{1}{2})^{m-\frac{1}{2}} e^{-m+\frac{1}{2}} = O(1)\Gamma(m+\tfrac{1}{2}).$$

Since

$$\Gamma(m+\tfrac{1}{2})/(2m+1)\,! = \pi^{\frac{1}{2}} 2^{-2m-1}/m\,!(m+\tfrac{1}{2}),$$

---

[3] G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. 1 (1925), p. 37, where it is proved for the interval (0, 1). It is easily extended to the interval (0, ∞).

it follows that the sum on the left side of (5) is asymptotic to

$$
\begin{aligned}
&2(6u/\pi)^{\frac{1}{2}} \sum_{m=1}^{\infty} (\pi^2/6u)^{m+1}/m!\,(m+\tfrac{1}{2}) \\
&= 2(6u/\pi)^{\frac{1}{2}} \sum_{m=1}^{\infty} \left((\pi^2/6u)^{m+1}/(m+1)! \right) (1 + O(1/m)) \\
&\sim 2(6u/\pi)^{\frac{1}{2}} \exp(\pi^2/6w).
\end{aligned}
$$

Proof of (6). Since $\sum_{n=1}^{\infty} p(n)x^n = \sum_{m=1}^{\infty} (1 - x^m)^{-1}$, we have

$$
\log\left(\sum_{n=1}^{\infty} p(n)e^{-nu} \right) = \sum_{n=1}^{\infty} n^{-1}(e^{nu} - 1)^{-1}
$$

$$
= -\sum_{n=1}^{\infty} \left(n^{-2}u^{-1} - (2n)^{-1}e^{-nu} - n^{-1}(e^{nu} - 1)^{-1} \right) + \pi^2/6u + 2^{-1}\log(1 - e^{-u}),
$$

and the last term may be replaced by $2^{-1}\log u + o(1)$.

Once again we may apply the lemma. It is well-known[4] that

$$
\int_0^{\infty} \left(t^{-2} - (2t)^{-1}e^{-t}e^{-t} - t^{-1}(e^t - 1)^{-1} \right) dt = 2^{-1}\log 2\pi.
$$

Hence

$$
\log\left(\sum_{n=1}^{\infty} p(n)e^{-nu} \right) = 2^{-1}\log(u/2\pi) + \pi^2/6u + o(1),
$$

which implies the desired result.

NEW YORK UNIVERSITY.

---

[4] E. T. Whittaker and G. N. Watson, *Modern Analysis*, American Edition, New York, 1944. On p. 249, line 15, let z tend to zero and apply p. 248, line 12.

UNIFORM CAUCHY POINTS AND POINTS OF EQUICONTINUITY.* [1]

By B. J. PETTIS.

Let X and Y be metric spaces with metrics ρ and σ, and let $\{f_n\}$ be a sequence of continuous functions on X to Y converging in X to a function f_0. A point x_0 is *a point of uniform convergence of* $\{f_n\}$ *to* f_0 if for each $\epsilon > 0$ there exist a $\delta_\epsilon > 0$ and a positive integer n_ϵ such that $\sigma(f_n(x), f_0(x)) < \epsilon$ holds whenever $\rho(x, x_0) < \delta_\epsilon$ and $n \geqq n_\epsilon$. Let U be the set of all such x_0. In 1897, antedating Baire's thesis, W. F. Osgood defined a set E in $[0, 1]$ to be a "P-set" [15, p. 161] if it is nowhere dense and "contains its derivative," i. e., E is nowhere dense and closed, and called E a "Q-set" [15, p. 171] if it is the union of a monotone increasing sequence of P-sets, that is, if it is a first category F_σ-set. He proved that the complement $X - U$ of U is a Q-set when X is $[0, 1]$, Y the reals, and f_0 is continuous. Shortly afterwards other authors [6, pp. 1140-3] removed the continuity hypothesis on f_0 and showed in addition that f_0 is continuous at each point of U. In 1924 Kuratowski [10] extended this to the case in which X is separable and Y metric, and in 1930 Banach [1] eliminated Kuratowski's separability assumption. An excellent proof for topological X and metric Y can be found in Hausdorff's book [9, pp. 285-6].

In this note we wish to extend the Osgood theorem to the case in which $\{f_\lambda\}$ is a directed set of continuous functions, X is a topological space, Y is a uniform space, and no limit function f_0 is assumed; in place of the omitted condition it is assumed that the set C of Cauchy points of $\{f_\lambda\}$ is, in a certain prescribed sense, a "residual set" in X. The resulting conclusion is that the set C_u composed of uniform Cauchy points and the set C_a of points of almost equi-continuity also are residual in the same sense. (The case in which f_0 exists is covered in Corollary 1. 1; the latter thus constitutes a direct generaliaztion of Hausdorff's version of the Osgood theorem.) The proof of Theorem 1 combines that of Hausdorff with a device due to A. Weil.

---

* Received June 13, 1950; revised October 20, 1950.

[1] Presented to the American Mathematical Society September 6, 1948, under the title *On points of uniform convergence in topological spaces.* This paper was written under Contract N7-onr-434, Task Order III, Navy Department (the Office of Naval Research).

In Theorem 2 we consider the case in which C has the property that its complement is not residual; the conclusion is that C_u and C_a have the same property. Applying Theorems 1 and 2 to the case in which $\{f_\lambda\}$ is essentially denumerable and Y is pseudo-metric, it follows (Theorem 3) that (i) if the complement of C is first category so are the complements of C_u and C_a, and (ii) if C is second category so are C_u and C_a. Several well known results in the literature can be considered as applications of Theorem 3; some of these—the theorems of Lebesgue-Hahn-Saks and of Nikodym concerning convergent sequences of completely additive set functions, the uniform boundedness of convergent sequences of linear operations in Banach spaces, the theorem of Montgomery on the double continuity of the group product in certain groups with topologies, Yosida's abstraction of Banach's theorem on convergence almost everywhere, etc.—are derived below and in certain cases in more generality.

When $\{f_\lambda\}$ is a denumerable sequence and Y is metric and X is a complete metric space the conclusions of Theorem 3 concerning C are identical with recent results due[2] to A. Alexiewicz [20, p. 5], results from which the latter has drawn a variety of theorems including two of the applications of Theorem 3 that were mentioned above. The two papers differ in that each contains applications not in the other, and in that Theorems 1, 2, and 3 of the present paper constitute generalizations of Alexiewicz's Theorem 1 and Corollary.

Let X be any topological space and ϕ any cardinal number. A subset E of X is a $I\phi$-set if it is the union of ϕ nowhere dense sets, and otherwise is a $II\phi$-set; it is ϕ-residual if its complement $X - E$ is a $I\phi$-set [5]. Let Y be any completely regular topological space, not necessarily Hausdorff, and of all uniformizations of the topology in Y [18] let $\{V_\alpha \mid \alpha \in A\}$ be one having least cardinal number $|A|$. For each element λ of a fixed directed set Λ [3] let f_λ be a function on X to Y. A point x_0 in X is *a Cauchy point of* $\{f_\lambda\}$ if (0.1) given any $\alpha \in A$ there is some $\lambda_\alpha \in \Lambda$ such that $f_\lambda(x_0) \in V_\alpha(f_{\lambda_\alpha}(x_0))$ for all $\lambda \geqq \lambda_\alpha$; and x_0 is *a uniform Cauchy point*[3] *of* $\{f_\lambda\}$ *if* (0.2) for each $\alpha \in A$ there exist a $\lambda_\alpha \in \Lambda$ and an open set G about x_0 such that $f_\lambda(x) \in V_\alpha(f_{\lambda_\alpha}(x))$ for all $\lambda \geqq \lambda_\alpha$ and all x in G. The set of Cauchy points will be denoted by C and the set of uniform Cauchy points by C_u. *A point*

---

[2] It was after submission of the present paper for publication that the author became aware of [20]. Results common to the two papers are clearly to the credit of Alexiewicz.

[3] A term suggested by the referee, in place of Du Bois-Reymond's *a point of uniform convergence* for a sequence $\{f_n\}$.

of almost equi-continuity of $\{f_\lambda\}$ is a point x_0 satisfying (0.3) for each $\alpha \varepsilon A$ there exist a $\lambda_\alpha \varepsilon \Lambda$ and an open G about x_0 such that $f_\lambda(x) \varepsilon V_\alpha(f_\lambda(x_0))$ for any $\lambda \geqq \lambda_\alpha$ and any x in G. The set of all such points will be denoted by C_a. If x_0 has the property that for each α there is an open G about x_0 such that $f_\lambda(x) \varepsilon V_\alpha(f_\lambda(x_0))$ for all x in G and all λ, we say that x_0 is *a point of equi-continuity of* $\{f_\lambda\}$, and denote by C_e the set of all such x_0. Clearly $C_e \subset C_a$, and if Λ is the set of positive integers and each f_λ is continuous then $C_e = C_a$. The closure and interior of any set E in a topological space will be represented by E^* and E^0 respectively.

The following lemma, which is well known when Y is metric and Λ is the set of positive integers [8], is easily verified in the present case.

LEMMA 1. *The relations* $C_u \supset C^* \cap C_a = C \cap C_a$ *always hold, and if* f_λ *is continuous for each* λ *in a cofinal subset* [3] *of* Λ *then* $C_u = C^* \cap C_a = C \cap C_a$.

Letting $|\Lambda|$ represent the least of the cardinal numbers belonging to cofinal subsets of Λ, our main result is

THEOREM 1. *Suppose that each* f_λ *is continuous, that* $\phi \geqq \max(|A|, |\Lambda|)$ *and that* C *is* ϕ*-residual. Then* C_u *and* C_a *are* ϕ*-residual and* $C_u \subset C_a$.

For each $\alpha \varepsilon A$ there exists [18, p. 14] a non-negative real function $\sigma_\alpha(p, q)$ on $Y \times Y$ such that (i) $\sigma_\alpha(p, p) = 0$, (ii) $\sigma_\alpha(p, q) = \sigma_\alpha(q, p)$, (iii) if $\sigma_\alpha(p, q) < 1$ then $p \varepsilon V_\alpha(q)$ and $q \varepsilon V_\alpha(p)$, and (iv) σ_α is uniformly continuous, i. e., given $\epsilon > 0$ there is some $\beta \varepsilon A$ such that $|\sigma_\alpha(r, s) - \sigma_\alpha(p, q)| < \epsilon$ whenever $r \varepsilon V_\beta(p)$ and $s \varepsilon V_\beta(q)$.

Let $M = [\mu]$ be a cofinal subset of Λ having cardinal number $|\Lambda|$, and for each $\alpha \varepsilon A$ and $\mu \varepsilon M$ let $Q_{\alpha\mu} = X[x \mid |\sigma_\alpha(f_\lambda(x), f_\mu(x))| \leqq \tfrac{1}{2}$ for all $\lambda \geqq \mu]$ and $R_{\alpha\mu} = X - Q_{\alpha\mu}$. The essential part of the proof is to establish, independent of the continuity of the f_λ's, that

$$C = \bigcap_\alpha \bigcup_\mu Q_{\alpha\mu} \text{ and } C_u = \bigcap_\alpha \bigcup_\mu Q^0{}_{\alpha\mu}.$$

The first of these implies that $X - C = \bigcup_\alpha \bigcap_\mu R_{\alpha\mu}$; since $X - C$ is a I_ϕ-set, so is $\bigcap_\mu R_{\alpha\mu}$ for each α. From the second equality, $X - C_u = \bigcup_\alpha \bigcap_\mu (X - Q^0{}_{\alpha\mu}) = \bigcup_\alpha \bigcap_\mu R^*{}_{\alpha\mu}$; if $\bigcap_\mu R^*{}_{\alpha\mu}$ were a I_ϕ-set for each α then, since $\phi \geqq |A|$, clearly $X - C_u$ must be a I_ϕ-set, i. e., C_u is ϕ-residual, and this combined with Lemma 1 provides the desired conclusion. Now $\bigcap_\mu R^*{}_{\alpha\mu} \subset [\bigcap_\mu R_{\alpha\mu}] \cup [\bigcup_\mu (R^*{}_{\alpha\mu} - R_{\alpha\mu})]$, where $\bigcap_\mu R_{\alpha\mu}$ has been noted to be a I_ϕ-set; since $\phi \geqq |\Lambda|$, $\bigcap_\mu R^*{}_{\alpha\mu}$ will then be a I_ϕ-set if each $R^*{}_{\alpha\mu} - R_{\alpha\mu}$ is a I_ϕ-set. But for fixed λ and μ the function $\sigma_\alpha(f_\lambda(x), f_\mu(x))$ is continuous in x by (iv) above, since f_λ and f_μ are continuous; thus each $Q_{\alpha\mu}$ is closed, each $R_{\alpha\mu}$ is open, and hence each $R^*_{\alpha\mu} - R_{\alpha\mu}$ is nowhere dense and therefore a I_ϕ-set.

To verify the equality $C = \bigcap_\alpha \bigcup_\mu Q_{\alpha\mu}$, first suppose $x_0 \varepsilon C$ and α is fixed. By (iv) and (i) above there is a $\beta \varepsilon A$ such that $r \varepsilon V_\beta(p)$ and $s \varepsilon V_\beta(p)$ imply $|\sigma_\alpha(r, s)| < \frac{1}{2}$. Having $\beta \varepsilon A$, $x_0 \varepsilon C$, and M cofinal in Λ, there is some $\mu_\beta \varepsilon M$ such that $f_\lambda(x_0) \varepsilon V_\beta(f_{\mu_\beta}(x_0))$ for all $\lambda \geqq \mu_\beta$. Since $f_{\mu_\beta}(x_0) \varepsilon V_\beta(f_{\mu_\beta}(x_0))$, it follows that $|\sigma_\alpha(f_\lambda(x_0), f_{\mu_\beta}(x_0))| < \frac{1}{2}$ for $\lambda \geqq \mu_\beta$, and hence $x_0 \varepsilon Q_{\alpha\mu_\beta}$. Thus $C \subset \bigcap_\alpha \bigcup_\mu Q_{\alpha\mu}$. Conversely, if $x_0 \varepsilon \bigcap_\alpha \bigcup_\mu Q_{\alpha\mu}$ then $x_0 \varepsilon Q_{\alpha\mu_\alpha}$ for each α, i. e., $|\sigma_\alpha(f_\lambda(x_0), f_{\mu_\alpha}(x_0))| \leqq \frac{1}{2}$ for all $\lambda \geqq \mu_\alpha$, so that by (iii) we have $f_\lambda(x_0) \varepsilon V_\alpha(f_{\mu_\alpha}(x_0))$ for all such λ. Thus $x_0 \varepsilon C$.

To establish the second equality, fix x_0. If for each α we have $x_0 \varepsilon Q^0{}_{\alpha\mu_\alpha}$ for some $\mu_\alpha \varepsilon M$ then, letting $G = Q^0{}_{\alpha\mu_\alpha}$, clearly $|\sigma_\alpha(f_\lambda(x), f_{\mu_\alpha}(x))| \leqq \frac{1}{2}$ holds for $\lambda \geqq \mu_\alpha$ and $x \varepsilon G$, so that by (iii) we have $f_\lambda(x) \varepsilon V_\alpha(f_{\mu_\alpha}(x))$ for such λ and x. Since this holds for each α, the point x_0 is in C_u. On the other hand, when x_0 is in C_u consider any fixed α. By (iv) and (i) there is some $\beta \varepsilon A$ such that $|\sigma_\alpha(r, s)| < \frac{1}{2}$ whenever $r \varepsilon V_\beta(p)$ and $s \varepsilon V_\beta(p)$. Having x_0 in C_u and M cofinal in Λ, there is some open G about x_0 and some $\mu_\beta \varepsilon M$ such that $f_\lambda(x) \varepsilon V_\beta(f_{\mu_\beta}(x))$ whenever $x \varepsilon G$ and $\lambda \geqq \mu_\beta$. For such x and λ we then have $|\sigma_\alpha(f_\lambda(x), f_{\mu_\beta}(x))| < \frac{1}{2}$, implying that $G \subset Q_{\alpha\mu_\beta}$ and hence $x_0 \varepsilon Q^0{}_{\alpha\mu_\beta}$, ending the proof.

It may be remarked at this point that the above proof that C_u is ϕ-residual still holds in case there is a secondary topology in X in which every open set has, in terms of the original topology, a I_ϕ-set for its boundary; if under these circumstances each f_λ is assumed continuous in the secondary topology rather than in the primary one the proof goes through with little change, the conclusions still being in terms of the primary topology.

Given any function f_0 on X to Y, let C_0 be the set of points of continuity of f_0; and given f_0 and $\{f_\lambda\}$, let $C(f_0)$ and $C_u(f_0)$ represent the sets of those points x_0 for which (0.1) and (0.2) respectively hold when f_{λ_α} is replaced by f_0. The verification of the following is similar to that of Lemma 1 and is also omitted.

LEMMA 2. *For any f_0 and any $\{f_\lambda\}$ these hold*: (1) $C \supset C(f_0)$; (2) $C_0 \supset C(f_0)^\circ \cap C_a$; (3) $C(f_0) \cap C_u \supset C_u(f_0) \supset C(f_0)^\circ \cap C_u$; (4) *if f_λ is continuous for each λ in a cofinal set then* $C_0 \supset C_u(f_0)$.

The following form of the Osgood theorem can now be derived.

COROLLARY 1. 1. *Given $\{f_\lambda\}$ and f_0, suppose that each f_λ is continuous.*

that $\phi \geqq \max(|A|, |\Lambda|)$, *and that* $C(f_0)^0$ *is* ϕ*-residual. Then* $C_u(f_0)$, $C_a \cap C_0$, *and* C_0 *are* ϕ*-residual, and* $C_u(f_0) \subset C_a \cap C_0$.

If $C(f_0)^0$ is ϕ-residual so is C by (1) of Lemma 2; from Theorem 1, C_u must be ϕ-residual. The intersection $C(f_0)^0 \cap C_u$ then has the same property; and since the lemmas give us $C(f_0)^0 \cap C_u \subset C_u(f_0) \subset C_u \subset C_a$ and $C_u(f_0) \subset C_0$, the corollary follows.

THEOREM 2. *Suppose that each* f_λ *is continuous, that* $\phi \geqq \max(|A|, |\Lambda|)$, *and that* C *is a* II_ϕ*-set. Then* $C \cap C_a$ $(= C_u)$ *is a* II_ϕ*-set.*

Let $g_\lambda(x) = f_\lambda(x)$ for $x \varepsilon C$. By Theorem 1 there is a set D which is ϕ-residual in the space C and has each of its points a point of almost equi-continuity of $\{g_\lambda\}$ in C. From its first property D is a II_ϕ-set in X since C is, and hence [5, Th. 3. 3] there exists a non-null open set G_D in X such that $N \cap G_D \cap D$ is a II_ϕ-set in X whenever N is open in X and $N \cap G_D$ is non-null. Letting $T = G_D \cap D$, T is a II_ϕ-set and $T \subset C$; since $C \cap C_a = C_u$ by Lemma 1, we have only to prove that $T \subset C_a$.

Fix $t \varepsilon T$ and $\alpha \varepsilon A$, and choose β so that if p, q, $r \varepsilon Y$, $p \varepsilon V_\beta(q)$, and $p \varepsilon V_\beta(r)$ then $q \varepsilon V_\alpha(r)$. Now $t \varepsilon D$ implies that for some neighborhood $N(t)$ of t in X and some $\lambda_\beta \varepsilon \Lambda$ we have $g_\lambda(x') \varepsilon V_\beta(g_\lambda(t))$ for any $\lambda \geqq \lambda_\beta$ and any $x' \varepsilon N(t) \cap C$. If it can be established that $f_\lambda(x) \varepsilon V_\alpha(f_\lambda(t))$ for any $\lambda \geqq \lambda_\beta$ and any x in the set $N(t) \cap G_D$, which is open in X and contains t, clearly t is in C_a. Fixing x and λ accordingly, the continuity of f_λ means that $f_\lambda(x') \varepsilon V_\beta(f_\lambda(x))$ for any x' in some set $N(x)$ open in X about x. Now $N(x) \cap N(t) \cap G_D$ is open and is non-null since it contains x, and hence $N(x) \cap N(t) \cap T$ is a II_ϕ-set and therefore non-null. Choose any x' in this set. Since $x' \varepsilon N(x)$ we have $f_\lambda(x') \varepsilon V_\beta(f_\lambda(x))$; and since $x' \varepsilon N(t) \cap T \subset N(t) \cap C$, a remark above yields the inclusion $g_\lambda(x') \varepsilon V_\beta(g_\lambda(t))$, whence $f_\lambda(x') \varepsilon V_\beta(f_\lambda(t))$. From the manner in which β was chosen we conclude that $f_\lambda(x) \varepsilon V_\alpha(f_\lambda(t))$, completing the proof.

Defining $\{f_\lambda\}$ to be *of Class* 1-ϕ if C_a is either a I_ϕ-set or the whole space X, there is

COROLLARY 2. 1. *Suppose in addition to the hypotheses of Theorem* 2 *that* $\{f_\lambda\}$ *is of Class* 1-ϕ. *Then* $C_a = X$ *and the sets* C_u, C, C^* *coincide and contain an open* II_ϕ*-set.*

From the theorem C_a is a II_ϕ-set and so $C_a = X$; from Lemma 1 we then have $C_u = C^* = C$. Since C is a closed II_ϕ-set its interior C^0 is a II_ϕ-set.

COROLLARY 2. 2. *Suppose in addition to the hypotheses of Corollary* 2. 1 *that* X *is a topological group and* C *is a subgroup of* X. *Then* $C_a = X$, *and* C *is an open and closed* II_ϕ*-subgroup; if* X *is connected then* $C = C_a = C_u = X$ *and if also* Y *is complete* [18] *then* $\lim_\lambda f_\lambda(x)$ *exists everywhere in* X *and is continuous.*

Corollary 2. 2 is an extension of a theorem of Banach's [2].

Calling Λ *essentially denumerable* if $|\Lambda| \leqq \aleph_0$ (e. g., $\Lambda =$ positive integers, $\Lambda =$ positive reals) and labeling a topological space Y *pseudo-metric* if the topology is given by a function σ having the properties of a metric except possibly that of $\sigma(y_1, y_2)$ being zero implying $y_1 = y_2$, the following is evident, the uniform topology in Y being that determined by σ. This result (in slightly less general form) is due to Alexiewicz [20].

THEOREM 3. *Let* Y *be pseudo-metric and* Λ *essentially denumerable and suppose* f_λ *is continuous for each* $\lambda \varepsilon \Lambda$. *Then* (1) *if* $X - C$ *is first category so are* $X - C_a$ *and* $X - C_u$; (2) *if* C *is second category so are* C_a *and* C_u; (3) *if* f_0 *is on* X *to* Y *and* $X - C(f_0)^0$ *is first category so are* $X - C_u(f_0)$, $X - C_a$, *and* $X - C_0$. *Hence if* X *is second category so are* C_u *and* C_a *in all three cases, and in the case of* (3) *so is* $C_u(f_0) \cap C_a \cap C_0$.

The last conclusion in Theorem 3 holds when X contains a conditionally compact non-null open set or when X is a locally complete pseudo-metric space.

Remarks on topological groups. The above results, being essentially double limit theorems, may be expected to have applications of the following kinds. The first is an extension of a well known result due to D. Montgomery [13].

THEOREM 4. *Let* X *be a group and also a space with a uniform structure, and suppose that* xy *is continuous in each variable separately. Suppose that* ϕ *is the least cardinal number among those belonging to families of neghborhoods fundamental in the uniform structure. If* X *is a* II_ϕ*-set then* xy *is continuous in* (x, y).

If xy is not continuous at a point (a, b) in $X \times X$ it is not continuous at (e, e) where e is the identity element. Let $\{V_\lambda \mid \lambda \varepsilon \Lambda\}$ be a fundamental family of neighborhoods having cardinal number ϕ. Since xy is not continuous at (e, e) there is an open set N_0 containing e and points x_λ, y_λ in each $V_\lambda(e)$ such that $x_\lambda y_\lambda$ is not in N_0. For each λ let $f_\lambda(x) = xy_\lambda$; each f_λ is continuous in x. Partially ordering Λ by defining $\mu \geqq \lambda$ to mean

$V_\mu(e) \subset V_\lambda(e)$, Λ is a directed set and, letting f_0 be the identity map in X, we have $C(f_0) = X$ since xy is continuous in y. Applying Corollary 1.1, $C_u(f)$ is ϕ-residual; since X is a II_ϕ-set $C_u(f_0)$ must be non-null. Choose $x_0 \varepsilon C_u$, pick $\beta \varepsilon \Lambda$ so that $x_0^{-1}V_\beta(x_0) \subset N_0$, and let γ in Λ be such that $p \varepsilon V_\gamma(q)$ and $q \varepsilon V_\gamma(r)$ imply $p \varepsilon V_\beta(r)$. Since x_0 is in $C_u(f_0)$, given γ there are δ and ϵ in Λ such that $xy_\lambda \varepsilon V_\gamma(x)$ whenever $\lambda \geqq \delta$ and $x \varepsilon V_\epsilon(x_0)$. Choosing first $\zeta \geqq \gamma, \epsilon$ and then η such that $V_\eta(e) \subset x_0^{-1}V_\zeta(x_0)$, clearly $\lambda \geqq \eta$ implies $x_0x_\lambda \varepsilon V_\zeta(x_0) \subset V_\gamma(x_0) \cap V_\epsilon(x_0)$. Now take any fixed $\mu \geqq \delta, \eta$. Since $\mu \geqq \eta$, it follows that $x_0x_\mu \varepsilon V_\epsilon(x_0)$ and hence, since $\mu \geqq \delta, \eta$, that $x_0x_\mu y_\mu \varepsilon V_\gamma(x_0x_\mu)$. But $\mu \geqq \eta$ also gives us $x_0x_\mu \varepsilon V_\gamma(x_0)$. These two inclusions imply that $x_0x_\mu y_\mu \varepsilon V_\beta(x_0)$, so that $x_\mu y_\mu \varepsilon x_0^{-1}V_\beta(x_0) \subset N_0$, which is contrary to the way in which x_μ and y_μ were originally chosen.

(2) For the second application let X be a topological group and let G be the group of all single-valued functions g on X to X with the group operation $(gh)(x) = g(x)h(x)$. Let H be a subgroup of G possessing a pseudo-metric ν with these properties: (i) ν is right invariant, i.e., $\nu(g, h) = \nu(g\bar{h}, h\bar{h})$ for any $g, h, \bar{h}$ in H, (ii) H is second category with respect to ν, (iii) $\lim_n \nu(h_n, h_0) = 0$ implies $\lim_n h_n(x) = h_0(x)$ for every x in X. Suppose also that given any nucleus U (open set about the identity e) in X there is a right invariant pseudo-metric ρ in X, depending on U, such that (iv) each ρ-sphere $S_\rho(a;\epsilon) \equiv X[x \mid \rho(x,a) < \epsilon]$ is open in X, (v) $S_\rho(e;1) \subset U$, and (vi) each h in H is continuous on X to X with respect to the ρ-topology. For one example, let X be a real or complex linear space with a topology such that X is a topological group satisfying the first countability axiom and scalar multiplication hx is continuous in h and x separately; by the Birkhoff-Kakutani metrization theorem there is a right invariant pseudo-metric $\bar{\rho}$ giving the group topology, and if H is taken to be all functions hx, h a scalar, and is given the scalar topology then (i)-(vi) are all satisfied when for a given U we take ρ to be $n\bar{\rho}$ with n chosen so that $S_{\bar{\rho}}(e;1/n) \subset U$. For another, let X be a real linear space with a topology such that X is a topological group, X has arbitrarily small non-null open convex sets about any point, and hx is continuous in each separate variable; let H again be the scalar multiples, and given U choose a convex nucleus K in U and let ρ be the pseudo-metric determined by the Minkowski functional of K. The next theorem, when applied to the first example, yields a result of Mazur and Orlicz [12, p. 188].

THEOREM 5. *Under the above assumptions on X and H the function $h(x)$ on $H \times X$ to X is continuous.*

If it were not continuous there would exist $x_0 \varepsilon X$, $h_0 \varepsilon H$, and an open set G about $h_0(x_0)$ such that for no choice of N open about x_0 and of $\delta > 0$ do we have $h(x) \varepsilon G$ for all x in N and all h with $\nu(h, h_0) < \delta$. Writing $\bar{x}$ for $h_0(x_0)$ and $U\bar{x}$ for G where U is a nucleus in X, choose ρ so that (iv)-(vi) hold. From (v) and the right invariance of ρ we have $S_\rho(\bar{x}; 1) \subset U\bar{x} = G$, and from (iv) $S_\rho(x_0; 1/n)$ is open about x_0 for $n = 1, 2, \cdots$. Hence there exist $x_n \varepsilon X$ and $h_n \varepsilon H$, $n = 1, 2, \cdots$, such that $\rho(x_n, x_0) < 1/n$, $\nu(h_n, h_0) < 1/n$, and $h_n(x_n)$ is not in G, i. e., $\rho(h_n(x_n), h_0(x_0)) \geqq 1$. Now let $f_n(h) = h(x_n)$. Condition (iii) obviously implies that each f_n is continuous on H to X; and for each h we have, by (vi), that $f_n(h) \to h(x_0)$ in the ρ-topology since $\rho(x_n, x_0) \to 0$. Applying (2) of Theorem 3 there exists a point $\bar{h}$ of almost equi-continuity of $\{f_n\}$. Hence there is a $\delta > 0$ such that $\rho(h(x_n), \bar{h}(x_n)) < \frac{1}{2}$ whenever n is sufficiently large and $\nu(h, \bar{h}) < \delta$. Now $\nu(h_n, h_0) < \delta$ for sufficiently large n and hence, using (i), $\nu(h_n h_0^{-1}\bar{h}, \bar{h}) < \delta$; it then follows that for large n we have $\frac{1}{2} > \rho(h_n(x_n) h_0(x_n)^{-1}\bar{h}(x_n), \bar{h}(x_n))$ $= \rho(h_n(x_n), h_0(x_n))$. But $\rho(x_n, x_0) \to 0$ and condition (vi) together imply that $\rho(h_0(x_n), h_0(x_0)) < \frac{1}{2}$ for large n. Hence $1 > \rho(h_n(x_n), h_0(x_0))$ for large n, which contradicts the choice of $\{x_n\}$ and $\{h_n\}$.

(3) In this section *X and Y are topological groups, Λ is essentially denumerable, and $\{h_\lambda \mid \lambda \varepsilon \Lambda\}$ is a family of continuous homomorphisms on X to Y.* The sets C, C_u, C_a, and C_e in X determined by the right uniform topology in Y will be denoted by $C(R)$, $C_u(R)$, etc., and those determined by the left uniform topology by $C(L)$, $C_u(L)$, etc. For any family of homomorphisms the sets $C_a(R)$ and $C_a(L)$ have these properties: (i) each is either the null set or the whole space, (ii) each contains e if and only if the other does. Hence $C_a(R) = C_a(L)$ and either both are null or both are the whole space. (This implies that any such family is of Class 1-ϕ for any ϕ and for any one of the two uniform topologies considered in Y; the next theorem may thus be considered as a variation of Corollary 2. 2.)

THEOREM 6. *Suppose that either $C(R)$ or $C(L)$ is second category. Then $C_a(R) = C_a(L) = X$, $C_u(R) = C(R) = C(R)^*$, and $C_u(L) = C(L)$ $= C(L)^*$; if Λ is the positive integers then also $C_e(R) = C_e(L) = X$.*

It is sufficient to show that for any nucleus V in Y a nucleus U can be found such that $h_\lambda(x) \varepsilon V$ holds for $x \varepsilon U$ and for sufficiently large λ. Suppose that $C(R)$ is second category. Given V, let σ be a right-invariant pseudometric such that $S_\sigma(e; 1) \subset V$ and each σ-sphere $S_\sigma(b; \epsilon)$ is open in the Y-topology [16]. The latter fact implies, using the continuity of h_λ, that h_λ is continuous on X to $[Y, \sigma]$. It also implies that for any $\epsilon > 0$ the set

$S_\sigma(e;\epsilon)y$ is one of the uniform neighborhoods in the right uniform topology in Y; hence for any x in C there is some λ_ϵ such that $h_\lambda(x) \varepsilon S_\sigma(e;\epsilon)h_{\lambda_\epsilon}(x)$ for all $\lambda \geqq \lambda_\epsilon$. By the right invariance of σ we have $h_\lambda(x) \varepsilon S_\sigma(h_{\lambda_\epsilon}(x);\epsilon)$ for $\lambda \geqq \lambda_\epsilon$, so that $\{h_\lambda(x)\}$ is a Cauchy sequence in the σ-topology for each $x \varepsilon C$. By Theorem 3 there is a point a in X, a nucleus U, and some $\lambda_1 \varepsilon \Lambda$ such that $\sigma(h_\lambda(x), h_\lambda(a)) < 1$ for all $x \varepsilon Ua$ and all $\lambda \geqq \lambda_1$. Using the right invariance of σ again it follows that $\sigma(h_\lambda(x'), e) < 1$ (and hence $h_\lambda(x') \varepsilon V$) for all $x' \varepsilon U$ and all $\lambda \geqq \lambda_1$. Thus $e \varepsilon C_a(R)$ and hence $C_a(R) = C_a(L) = X$. The rest of the theorem now follows from Lemma 1 and a remark in the introduction. The case in which $C(L)$ is second category follows by a dual argument.

When Λ is the positive integers and X and Y are spaces of Type (F) the theorem just completed was established by Mazur and Orlicz [11]. Together with (2) of Lemma 2, Theorem 6 yields the following, proved for Λ the positive integers and for complete metric X and metric Y by Banach [2, 16].

COROLLARY 6.1. *If $h_\lambda(x)$ converges to $h_0(x)$ for every x in X, where X is second category and Y is a Hausdorff group, h_0 is a continuous homomorphism.*

It is also obvious that when X and Y are linear normed spaces and Λ is the set of positive integers Theorem 6 reduces to the well known assertion that $\{h_\lambda\}$ is uniformly bounded.

(4) For the final application to groups let X and Y be additively written topological groups, with θ denoting the zero element in each. Suppose also that in Y there are defined a partial ordering and a pseudo-metric σ such that (i) σ gives the topology in Y, (ii) for any sequence $\{y_n\}$ of points in Y with $y_n \geqq y_m$ for $n \geqq m$ these are equivalent: (α) $\{y_n\}$ has an upper bound; (β) $\lim \sigma(y_m, y_n) = 0$ as $m, n \to \infty$; (γ) $\lim_n \sigma(y_n, y_0) = 0$ for some y_0 in Y. We also assume throughout this part that *X is connected, Λ is essentially denumerable*, and that *$\{f_\lambda \mid \lambda \varepsilon \Lambda\}$ is a family of continuous functions on X to Y with $f_\mu(x) \geqq f_\lambda(x)$ for all x in X whenever $\mu \geqq \lambda$.* For the uniform topology in Y we take that determined by σ.

For convenience, a family $\{f_\lambda\}$ of functions on X to Y will be said to be *of Class 2* if either $C \cap C_a$ is first category or $C_a = X$; for example, if σ is right or left invariant any family of homomorphisms is of Class 2, for then C_a being non-null implies $C_a = X$.

THEOREM 7. *Let $\{f_\lambda\}$ be of Class 2 and let $L = X[x \mid \{f_\lambda(x)\}$ has an* *upper bound*]. *If* L *contains a second category subgroup of* X *there exists a continuous function* f_0 *such that* $\lim_\lambda \sigma(f_\lambda(x), f_0(x)) = 0$ *for every* x *in* X.

Let G be a second category subgroup contained in L. Since G^* is a non-null open and closed subgroup and X is connected, G^* must be X and so $L^* = X$. From the equivalence of (α) and (β) in (ii) and the essential denumerability of Λ, $L = C$; hence $C^* = X$ and C is second category. Applying Theorem 2, $C \cap C_a$ is second category; this, with $\{f_\lambda\}$ being of Class 2, yields $C_a = X$. Thus $C^* \cap C_a = X$ and hence, by Lemma 1, $C = X$. This, with (ii), means that for each x there is some element $f_0(x)$ such that $\lim_\lambda \sigma(f_\lambda(x), f_0(x)) = 0$. Obviously $C(f_0) = C = X$, whence, by (2) of Lemma 2, $C_0 = X$, i. e., f_0 is continuous.

COROLLARY 7. 1. *Suppose each* f_λ *is a homomorphism,* L *is second category, and* σ *is a right* (*or left*) *invariant metric. Then there is a continuous homomorphism* f_0 *on* X *to* Y *such that* $\lim_\lambda \sigma(f_\lambda(x), f_0(x)) = 0$ *for all* x.

From a remark preceding Theorem 7, $\{f_\lambda\}$ is of Class 2; from the equivalence of (α) and (γ) in (ii) it is easy to see that L is a subgroup; the conclusion now results from Theorem 7 and Corollary 6. 1.

COROLLARY 7. 2. *Suppose that* Y *is a po-group* [3], *that* $\{f_\lambda\}$ *is of Class 2, and that* $f_\lambda(x_1 + x_2) \leqq f_\lambda(x_1) + f_\lambda(x_2)$ *and* $f_\lambda(x) = f_\lambda(-x)$ *for arbitrary* x_1, x_2, x, *and* λ. *If* L *is second category the conclusion of Theorem 7 follows.*

From the assumptions on $\{f_\lambda\}$ and the fact that Y is a po-group clearly L is a subgroup; the conclusion is therefore immediate.

Now suppose that X is a real (or complex) linear topological space (hence connected), and that Y is a similar space but with a partial ordering and a pseudo-metric σ that satisfy not only (i) and (ii) above but also these: (iii) Y is a po-group, (iv) σ is right invariant, and (v) $y_1 \geqq y_2 \geqq -y_1$ implies $\sigma(y_1, \theta) \geqq \sigma(y_2, \theta)$. (Let S be a measure space with the measure of S finite; the space (M) of essentially bounded measurable real functions with the Fréchet metric and the spaces L_p, $p \geqq 1$, are examples of such Y.) Setting $\| y \| = \sigma(y, \theta)$ we note that $\| y_1 - y_2 \| = \sigma(y_1, y_2)$, $\| y \| = \| -y \|$, and $\| y_1 + y_2 \| \leqq \| y_1 \| + \| y_2 \|$.

COROLLARY 7. 3. *Let* $\{f_\lambda\}$ *be such that* (1) $f_\lambda(x) \geqq \theta$ *for all* λ *and* x, (2) $f_\lambda(x_1 + x_2) \leqq f_\lambda(x_1) + f_\lambda(x_2)$, (3) $f_\lambda(x) = f_\lambda(-x)$, *and* (4)

$\| f_\lambda(x)/k \| = \| f_\lambda(x/k) \|$ *for* $\lambda \in \Lambda$, $x \in X$, *and any positive integer* k. *If* L *is second category the conclusion of Theorem 7 follows.*

From Corollary 7.2 it is sufficient to show that $\{f_\lambda\}$ is of Class 2. To do this we first observe that if $C_a \ni \theta$ then $C_a = X$. For clearly $f_\lambda(x) - f_\lambda(a) \leqq f_\lambda(x - a)$; and $f_\lambda(a) \leqq f_\lambda(a - x) + f_\lambda(x) = f_\lambda(x - a) + f_\lambda(x)$ whence $-f_\lambda(x - a) \leqq f_\lambda(x) - f_\lambda(a)$. From (v) we then have $\| f_\lambda(x) - f_\lambda(a) \| \leqq \| f_\lambda(x - a) \|$, and hence $C_a = X$ if $C_a \ni \theta$. To complete the proof it will be shown that $C_a \ni \theta$ when $C^* \cap C_a$ is non-null.

Let $\epsilon > 0$ be given and choose $\delta > 0$ so that $\| \alpha y \| < \epsilon/2$ whenever $\| y \| < \delta$ and α is a scalar with $| \alpha | < \delta$. Choose $a \in C^* \cap C_a$. Since $a \in C_a$, there is a nucleus U in X such that $\| f_\lambda(x + a) - f_\lambda(a) \| < \delta$ for x in U and λ sufficiently large. Thus

$$(*) \qquad \| \alpha f_\lambda(x + a) - \alpha f_\lambda(a) \| < \epsilon/2 \text{ for } x \in U,\ | \alpha | < \delta,\ \lambda \text{ large.}$$

Now choose $b \in C \cap (U + a)$. Since $b \in C$ implies $\| f_\lambda(b) - f_{\lambda_1}(b) \| < \delta$ for all $\lambda \geqq$ some λ_1, so that $\| \alpha f_\lambda(b) - \alpha f_{\lambda_1}(b) \| < \epsilon/2$ for $| \alpha | < \delta$ and $\lambda \geqq \lambda_1$, and since $\| \alpha f_{\lambda_1}(b) \| \to 0$ as $| \alpha | \to 0$, we can choose a positive integer k such that $1/k < \delta$ and $\| f_\lambda(b)/k \| < \epsilon$ for large λ. On the other hand, $b \in U + a$, $1/k < \delta$, and (*) give us $\| f_\lambda(b)/k - f_\lambda(a)/k \| < \epsilon/2$. Combining the last two inequalities furnishes $\| f_\lambda(a)/k \| < 2\epsilon$ for large λ. Using (*) again, we also have $\| f_\lambda(x + a)/k - f_\lambda(a)/k \| < \epsilon/2$ for x in U and λ large. At this point we note that $\| f_\lambda(x/k) \| \leqq \| f_\lambda(x + a)/k - f_\lambda(a)/k \| + 2 \| f_\lambda(a)/k \|$ holds for any x, a, k, and λ, as the reader can easily verify from hypotheses (3) and (4) and the properties of $\| y \|$. Hence $\| f_\lambda(x/k) \| < 5\epsilon$ is true for x in U and λ large. Since kx is continuous in x there is a nucleus U' such that $kU' \subset U$. For large λ and x' in U' we now have $\| f_\lambda(x') \| < 5\epsilon$, so that $\theta \in C_a$, ending the proof.

Corollary 7.3 can be considered as a slight extension of Yosida's abstraction [19, Theorem 1] of the Banach theorem on convergence almost everywhere as formulated and proved by Mazur and Orlicz [11]. In using Yosida's theorem or Corollary 7.3 to obtain any concrete individual ergodic theorem, as some authors have done, the essential difficulty is that of establishing that L is second category; this is done, for example, in the paper "On the ergodic theorem," *Transactions of the American Mathematical Society,* vol. 60 (1946), pp. 538-549, by Nelson Dunford and D. S. Miller in their Lemma 6.

Remarks on completely additive set functions. Let X be a σ-field of sets in some space S and $m(x)$ a finite non-negative completely additive real function on X; with the pseudo-metric $\rho(x_1, x_2) = m((x_1 - x_2) \cup (x_2 - x_1))$, X is a complete pseudo-metric space. Let G be an additive abelian group having an invariant pseudo-metric and let $F(m)$ be the space of functions defined and additive on X to G and absolutely continuous with respect to m (i. e., continuous with respect to ρ at the null set). Each f in $F(m)$ is completely additive. Moreover, as is easily verified, the set C_e for any subset H of $F(m)$ is either vacuous or all of X; in particular, taking H to consist of an arbitrary single element f, C_e for f is the set of continuity points of f, which shows that each f in $F(m)$ is continuous with respect to ρ. Let $\{f_n\}$ be any countable sequence in $F(m)$, so that $C_a = C_e$. If C for $\{f_n\}$ is second category, Theorem 3 establishes that C_a is second category; hence $X = C_a = C_e$, which is the Lebesgue-Hahn-Saks theorem in the form given by Saks [17]. In particular, if $C = X$ then $C_e = X$; if moreover G is complete in its pseudo-metric then $\lim_n f_n(x)$ exists on X to G and by (3) of Theorem 3 must be in $F(m)$, establishing another result given by Saks [17]. (Abstractions of Saks' theorems have been proved in this manner by Alexiewicz [20].)

Given the σ-field X and any denumerable sequence $\{f_n\}$ of completely additive finite real (or complex) functions on X, let ν_n be the total variation function of f_n, let $k_n = 1/2^n(1 + \nu_n(S))$, and set $\bar{m}(x) = \sum_i^\infty k_n \nu_n(x)$; it is easily seen that $\bar{m}$ is an m of the type above and that $\{f_n\} \subset F(\bar{m})$. If the set C for $\{f_n\}$ is X then $C_e = C_a = X$ by what was just proved; if we set $f_0(x) = \lim_n f_n(x)$ we then have, by (2) of Lemma 2, that $C_0 \supset C(f_0)^0 \cap C_e = X$, and hence f_0 is continuous on X. Being obviously additive f_0 is then in $F(\bar{m})$ and thus is completely additive. This result is due to Nikodym [14] and the present proof to Saks [17]. We may also note in this case that since $C_a = C_e = X$ the null set is an element of C_e and hence $f_n(x) \to 0$ uniformly in n as $\bar{m}(x) \to 0$; if $\{x_k\}$ is any denumerable disjoint sequence in X it is now clear that $\lim \sum_{i=k}^\infty f_n(x_i) = 0$ uniformly in n as $k \to \infty$. This is a recent theorem of Doubrovsky [4].

Reverting to groups we make a final remark that a recent result of Gottschalk's [7] can be given the following form, using Theorems 3 and 4 and part of Gottschalk's proof. Let X be a group and also a locally complete pseudo-metric space. Suppose the center of X is dense in X and the group product xy is continuous in x for each y. Then xy is continuous in (x, y).

TULANE UNIVERSITY.

REFERENCES.

[1] S. Banach, "Théorème sur les ensembles de première catégorie," *Fundamenta Mathematicae*, vol. 16 (1930), pp. 395-398.

[2] ———, "Über metrische Gruppen," *Studia Mathematica*, vol. 3 (1931), pp. 101-113.

[3] G. Birkhoff, *Lattice Theory* (Rev. Ed.), American Mathematical Society Colloquium Publications, vol. 25, New York, 1948.

[4] V. Doubrovsky, "On some properties of completely additive set functions etc., *Recueil Mathématique*, New Series, vol. 20(62) (1947), pp. 317-329.

[5] N. Dunford, "On continuous mapping," *Annals of Mathematics*, vol. 41 (1940), pp. 639-661.

[6] *Encyclopädie der Mathematischen Wissenschaften*, vol. II, 3, II, Leipzig, 1923-7.

[7] W. H. Gottschalk, "Transitivity and equi-continuity," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 982-984.

[8] H. Hahn, *Reelle Funktionen*, Leipzig, 1932.

[9] F. Hausdorff, *Mengenlehre*, 3. Auflage, Berlin and Leipzig, 1935.

[10] C. Kuratowski, "Sur les fonctions représentables analytiquement et les ensembles de première catégorie," *Fundamenta Mathematicae*, vol. 5 (1924), pp. 75-86.

[11] S. Mazur and W. Orlicz, "Über Folgen linearer Operationen," *Studia Mathematica*, vol. 4 (1933), pp. 152-157.

[12] ———, "Sur les espaces métriques linéaires (I)," *Studia Mathematica*, vol. 10 (1948), pp. 184-208.

[13] D. Montgomery, "Continuity in topological groups," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 879-882.

[14] O. Nikodym, "Sur les suites convergences de fonctions parfaitement additives d'ensemble abstrait," *Monatshefte für Mathematik und Physik*, vol. 40 (1933), pp. 965-970.

[15] W. F. Osgood, "Non-uniform convergence and the integration of series term by term," *American Journal of Mathematics*, vol. 19 (1897), pp. 155-190.

[16] B. J. Pettis, "On the continuity and openness of homomorphisms in topological groups," *Annals of Mathematics*, vol. 52 (1950), pp. 293-308.

[17] S. Saks, "Addition to the note 'On some functionals'," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 965-970.

[18] A. Weil, "Sur les Espaces à Structure Uniforme et sur la Topologie Générale," *Actualités Scientifiques et Industrielles*, no. 551, Paris, 1937.

[19] K. Yosida, "An abstract treatment of the individual ergodic theorem," *Proceedings of the Imperial Academy (Tokyo)*, vol. 16 (1940), pp. 280-284.

[20] A. Alexiewicz, "On sequences of operations (I)," *Studia Mathematica*, vol. 11 (1949), pp. 1-30.

AN ITERATION FORMULA FOR FREDHOLM INTEGRAL EQUATIONS OF THE FIRST KIND.*

By L. LANDWEBER.

1. Introduction. Neumann's Method of solving Fredholm integral equations of the *second* kind by iteration is of great practical and theoretical value. For Fredholm integral equations of the *first* kind, on the other hand, Hellinger and Toeplitz [1] remark that a method of solution by iteration is not available.

Physical problems often lead to an integral equation of the first kind to which a good first approximation may be derived by physical reasoning. An example of this is the problem of determining an axial source-sink or doublet distribution which would yield the axially-symmetric potential flow about a body of revolution in a uniform stream. This problem leads to an integral equation of the *first* kind, $\frac{1}{2} = \int_0^1 [(x-t)^2 + y(x)^2]^{-3/2} m(t)\,dt$, where the axis of the body coincides with the x-axis from $x = 0$ to $x = 1$, $y(x)$ is a known function, representing the ordinates of the intersection of the given surface with a meridian plane and $m(x)$ is an unknown function, representing the distribution of the doublet strength per unit length along the axis. A well-known, excellent, first approximation to the doublet distribution for elongated bodies of revolution is [2] $m_0(x) = [y(x)]^2/4$. In cases such as this it would be highly desirable to have a method of successive approximations for improving upon this approximation.

The theories of Schmidt and Picard furnish expressions for solutions to integral equations of the first kind. However, these expressions are of little practical value since they involve the characteristic numbers and functions of an arbitrary kernel, and the methods for obtaining these are both tedious and approximate.

It is proposed to present an iteration formula for obtaining successive approximations to the solution of Fredholm integral equations of the *first* kind, and to prove the convergence of the successive approximations under various conditions.

---

* Received August 10, 1950.

2. The integral equation of the first kind; theory of Schmidt and Picard. We are concerned with solutions and approximations to solutions of the integral equation of the first kind

$$f(x) = \int_a^b k(x, y) g(y) dy, \tag{1}$$

where $f(x)$ and $k(x, y)$ are given continuous real functions in $a \leqq x, y \leqq b$, and $g(y)$ is an unknown function. As is well known, (1) may be transformed into the integral equation with a symmetric kernel,

$$F(x) = \int_a^b K(x, y) g(y) dy, \tag{2}$$

where

$$K(x, y) = \int_a^b k(t, x) k(t, y) dt \tag{3}$$

and

$$F(x) = \int_a^b k(y, x) f(y) dy. \tag{4}$$

A theory due to E. Schmidt [3] shows that there exists a set $\{\lambda_i\}$ of positive characteristic numbers, which may be supposed arranged in increasing order of magnitude, and corresponding adjoint sets $\phi_i(x)$ and $\psi_i(x)$ of real, continuous, orthonormalized characteristic functions, $(i = 1, 2, \cdots)$, such that

$$\phi_i(x) = \lambda_i \int_a^b k(x, y) \psi_i(y) dy, \qquad \psi_i(x) = \lambda_i \int_a^b k(y, x) \phi_i(y) dy. \tag{5}$$

It will be convenient, hereafter, to employ the customary operator notation for integral transforms, viz., $kg \equiv \int_a^b k(x, y) g(y) dy$, $Kg \equiv \int_a^b K(x, y) g(y) dy$; furthermore, since the range of variation and the integration limits will always be from a to b, specific reference to these limits will be omitted and we will frequently write integrals in an abbreviated form, viz., $\int_a^b f(x)\phi_i(x) dx \equiv \int f\phi_i$.

If the kernel $k(x, y)$ is degenerate, the number of characteristic functions is finite and they can be found by a well known procedure [4]. If $f(x)$ is expressible in the form $f(x) = \sum_{i=1}^{n} a_i \phi_i(x)$, the solution of (1) is

$$g(x) = \sum_{i=1}^{n} \lambda_i a_i \psi_i(x), \qquad a_i = \int f\phi_i. \tag{6}$$

If $f(x)$ is not of the above form, then (6) gives the best approximate solution of (1) in the least square sense, as can easily be shown. If the kernel $k(x, y)$

is non-degenerate, the sets $\{\lambda_i\}$, $\{\phi_i(x)\}$ and $\{\psi_i(x)\}$ are infinite. Since the degenerate case is readily disposed of, only the non-degenerate case will be considered hereafter.

These characteristic numbers and adjoint functions have several properties which will be required in the following:

a) λ_i^2 and $\psi_i(x)$ are characteristic numbers and functions of $K(x, y)$ [5], i. e.,

$$\psi_i = \lambda_i^2 K \psi_i. \tag{7}$$

b) A positive lower bound for the set $\{\lambda_i\}$ is given by the inequality [3]

$$1/\lambda_1^2 < \int\int k^2(x, y)\, dx dy. \tag{8}$$

c) Expansion Theorems. Every function $f(x)$ of the form (1), where $g(y)$ is any piecewise-continuous function, can be expanded in the absolutely and uniformly convergent series [5]

$$f(x) = \sum_{i=1}^{\infty} a_i \phi_i(x); \qquad a_i = \int f\phi_i = (\int g\psi_i)/\lambda_i. \tag{9}$$

Every function $F(x)$ of the form (4), where $f(x)$ is any piecewise-continuous function, can be expanded in the absolutely and uniformly convergent series

$$F(x) = \sum_{i=1}^{\infty} c_i \psi_i(x); \qquad c_i = \int F\psi_i = (\int f\phi_i)/\lambda_i. \tag{10}$$

If f is the same function in (9) and (10), the relations between the "Fourier" coefficients may be written

$$c_i = \int F\psi_i = (\int f\phi_i)/\lambda_i = (\int g\psi_i)/\lambda_i^2. \tag{11}$$

In general a solution of (1) does not exist. A theorem due to E. Picard [6] states that, *if the orthogonal set* ϕ_i *is complete, a solution of the integral equation* (1) *exists if and only if the series*

$$\sum_{i=1}^{\infty} \lambda_i^2 a_i^2,\ a_i = \int f\phi_i \tag{12}$$

is convergent.

In the Schmidt-Picard theory, the solution of (1) is intimately related to the sequence

$$\bar{g}_n = \sum_{i=1}^{n} \lambda_i a_i \psi_i(x), \qquad n = 1, 2, \cdots \tag{13}$$

as is expressed in the following theorems:

THEOREM 1. *The sequence* $\{k\bar{g}_n\}$ *converges in the mean to* $f(x)$ *if and only if the set* $\{\phi_i\}$ *is complete relative to* $f(x)$. *The sequence converges uniformly to* $f(x)$, *if a piecewise-continuous solution of the integral equation* (1) *exists.*

THEOREM 2. *If a piecewise-continuous solution* $g(x)$ *of* (1) *exists, the sequence* $\{\bar{g}_n\}$ *converges in the mean to* $g(x)$ *if and only if the set* $\{\psi_i\}$ *is complete relative to* $g(x)$. *If* $g(x)$ *is of the form* $\int k(y,x)h(y)\,dy$, *where* $h(y)$ *is any piecewise-continuous function, then the sequence* $\bar{g}_n$ *converges uniformly to* $g(x)$.

The completeness conditions on the sequences $\{\phi_i\}$ and $\{\psi_i\}$ in Theorems 1 and 2 refer to the so-called completeness relations

$$\int f^2 = \sum_{i=1}^{\infty} a_i^2, \quad a_i = \int f\phi_i \quad \text{and} \quad \int g^2 = \sum_{i=1}^{\infty} b_i^2, \quad b_i = \int g\psi_i. \tag{14}$$

The phrase "complete relative to $f(x)$" in Theorem 1 signifies that (14) need be satisfied only by the particular function $f(x)$, a condition which is considerably weaker than the assumption that the set $\{\phi_i\}$ is complete relative to a class of functions. Similarly (14) is assumed to apply only to the particular function $g(x)$ in Theorem 2.

The first part of Theorem 1 is of especial interest since it indicates that with increasing n, the error due to the assumption of $\bar{g}_n(x)$ as an approximate solution of (1) diminishes in a least square sense, even if a solution of (1) does not exist. However the disagreeable possibility exists that, beyond some value of n, the error may accumulate and increase at some values of x. Nevertheless, even in this case, such a sequence may give useful successive approximations in a particular problem, if the errors are observed at each step, and the approximations stopped when the error exceeds an acceptable value at any point.

The second part of Theorem 1 asserts that, for sufficiently large n, $\bar{g}_n$ satisfies the integral equation (1) as closely as desired. It is noteworthy that no assumptions are made with regard to the convergence of the sequence $\{\bar{g}_n\}$. Indeed, Theorem 2 shows that an additional condition is necessary to assure even convergence in the mean.

The expression (13) for $\bar{g}_n$, however, is of little practical value since it is expressed in terms of the characteristic numbers and functions of the kernel $k(x,y)$. Principally for these reasons the Fredholm integral equation of the first kind has been considered to be of little value [7]. On the other hand, another readily calculable sequence of functions $\{g_n(x)\}$ will be defined, which, it will be shown, has properties relative to a solution of the integral equation (1) identical to those of $\bar{g}_n(x)$.

3. The iteration formula. Let us now extend the operator notation, denoting $K^r g \equiv \int \cdots \int K(x, y_r) K(y_r, y_{r-1}) \cdots K(y_2 y_1) g(y_1)\, dy_r dy_{r-1} \cdots dy_1$. This notation is appropriate since the relation $K^r(K^s g) \equiv K^{r+s} g$ is satisfied, as is easily verified.

Let $g_0(x)$ be an assumed, approximate, piecewise-continuous solution of the integral equation (1). Then a set of continuous functions $g_1(x), g_2(x), \cdots$ is defined by the iteration formula

$$g_n = g_{n-1} + F - K g_{n-1} \tag{15}$$

where K and F are the functions defined in equations (3) and (4). The convergence of this sequence of functions and the applicability of its members as successive approximations to a solution of the integral equation (1) is the subject of the subsequent discussion.

The recurrence formula (15) can be readily solved for g_n in terms of g_0. First put

$$\gamma_n = g_n - g_{n-1}. \tag{16}$$

Then

$$g_n = g_0 + \sum_{i=1}^{n} \gamma_i \tag{17}$$

and also (15) may be written as

$$\gamma_n = F - K g_{n-1}. \tag{18}$$

Thus the γ_n's are not only the differences between successive g_n's but also serve as measures of the errors corresponding to the g_n's as approximate solutions of the iterated integral equation (2). Now, from (18), we have $\gamma_n - \gamma_{n-1} = -K\gamma_{n-1}$ or, in operation notation, $\gamma_n = (1-K)\gamma_{n-1}$. Hence, since the operator K satisfies the associative laws of multiplication, we obtain

$$\gamma_n = (1-K)^{n-1} \gamma_1, \tag{19}$$

where $(1-K)^{n-1}$ is to be formally expanded by the binomial theorem before operating on γ_1. Substituting for the γ_1 in equation (17) from equation (19), and performing the indicated summation, we obtain

$$g_n = g_0 + \{[1-(i-K)^n]/K\}(F - K g_0), \tag{20}$$

where, in the fractional operator, $(1-K)^n$ is to be expanded by the binomial theorem and a factor K in the numerator cancelled with the denominator before operating on $(F - K g_0)$.

If the sequence $\{g_n(x)\}$ converges uniformly, it is clear from (15), that $\lim g_n$ is a solution of the iterated integral equation (2). However, since an integral equation of the first kind has a solution only under special circumstances, $\{g_n(x)\}$ may not converge uniformly, and indeed may not converge at all. Nevertheless the g_n's may serve as useful approximations to a solution of (1) and (2) as will be evident on the basis of the convergence theorems in the next section.

4. Convergence theorems. It will be assumed hereafter that

$$\int_a^b \int_a^b k^2(x,y)\,dxdy \leqq 2. \tag{21}$$

This is no restriction since $k(x,y)$ can always be modified, so as to satisfy (21), by multiplying (1) by a suitable factor and, in the right member of the equation, incorporating the factor into the kernel.

The convergence theorems will first be stated and discussed briefly before their proofs are presented.

THEOREM 3. *The sequence $\{Kg_n\}$ converges uniformly to $F(x)$.*

Theorem 3 is very strong. Without any restrictive assumptions about completeness, the existence of a solution, or the convergence of the sequence $\{g_n\}$, it asserts that, for sufficiently large n, g_n satisfies the iterated integral equation (2) as closely as desired. Basically, however, our interest is in the integral equation (1), rather than with (2). Concerning the suitability of the g_n's as approximate solutions of (1) we have the weaker theorems.

THEOREM 4. *The sequence $\{kg_n\}$ converges in the mean to $f(x)$ if and only if the set $\{\phi_i\}$ is complete relative to $f(x)$. The sequence converges uniformly to $f(x)$ if a piecewise-continuous solution of the integral equation* (1) *exists.*

It will now be supposed that the 0-th approximation $g_0(x)$ is chosen of the form

$$g_0(x) = \int k(y,x)h(y)\,dy, \tag{22}$$

where $h(y)$ is any piecewise-continuous function. The special case $h(y) \equiv 0$ is also allowed. Concerning the convergence of the sequence $\{g_n\}$ we then have

THEOREM 5. *If a piecewise-continuous solution $g(x)$ of* (1) *exists, the sequence $\{g_n\}$ converges in the mean to $g(x)$ if and only if the set $\{\psi_i\}$ is complete relative to $g(x)$. If $g(x)$ is of the form $\int k(y,x)h(y)\,dy$, where $h(y)$ is any piecewise-continuous function, then the sequence $\{g_n\}$ converges uniformly to $g(x)$.*

It should be noted that Theorems 4 and 5 are identical, word for word, with Theorems 1 and 2 except for the substitution of g_n for $\bar{g}_n$. Hence the remarks concerning the suitability of the $\bar{g}_n$'s as approximations to a solution of the integral equation (1) are applicable to the g_n's as well.

In order to prove the foregoing theorems it is first convenient to establish several lemmas. Put

$$(23) \qquad F_n(x) \equiv Kg_n, \qquad f_n(x) \equiv kg_n.$$

The "Fourier" coefficients of F_n, f_n and g_n then satisfy the relations

$$(24) \qquad c_{in} = \int F_n\psi_i = (\int f_n\phi_i)/\lambda_i = (\int g_n\psi_i)/\lambda_i^2.$$

We then have

LEMMA 1. *$F_n(x)$ and $f_n(x)$ can be expanded in the absolutely and uniformly convergent series*

$$(25) \quad F_n(x) = \sum_{i=1}^{n} c_{in}\psi_i(x), \quad f_n(x) = \sum_{i=1}^{n} \lambda_i c_{in}\phi_i(x), \qquad n = 0, 1, 2, \cdots.$$

If $g_0(x)$ is chosen of the form (22), then also $g_n(x)$ may be expanded in the absolutely and uniformly convergent series

$$(26) \qquad g_n(x) = \sum_{i=1}^{n} \lambda_i^2 c_{in}\psi_i(x), \qquad n = 0, 1, 2, \cdots.$$

Proof. It is clear, from their definitions in (23), that the expansion theorems apply to $F_n(x)$ and $f_n(x)$ and consequently the series (25) converge as stated in the lemma. In the case of the g_n's, it can be shown successively, from the iteration formula (15), that $g_1(x), g_2(x), \cdots$ are of the same form as $g_0(x)$. Thus we have

$$(27) \qquad g_1 = g_0 + F - Kg_0.$$

But $g_0 = \int k(y, x)h(y)dy$; from (4), $F = \int k(y, x)f(y)dy$; and from (3), (23), $Kg_0 = \int k(y, x)f_0(y)dy$. Hence (27) becomes $g_1 = \int k(y, x)[h(y) + f(y) - f_0(y)]dy$. Hence the expansion theorem is applicable to $g_n(x)$ and the series (26) also converge, as stated.

LEMMA 2.

$$(28) \qquad c_{in} - c_i = \mu_i^n(c_{i0} - c_i),$$

where $c_i = \int F\psi_i$, and the sequence μ_i is such that

$$(29) \qquad |\mu_i| < 1, \quad \mu_{i+1} \geqq \mu_i \text{ and } \lim_{i\to\infty} \mu_i = 1, \qquad i = 1, 2, \cdots.$$

Proof. We obtain, from (15) and (7), $\int g_n\psi_1 = (1 - 1/\lambda_i^2)\int g_{n-1}\psi_i$

$+ \int F\psi_i$. Put $\mu_i = 1 - 1/\lambda_i^2$. Then, by successive reduction, we obtain $\int g_n\psi_i = \mu_i^n \int g_0\psi_i + \lambda_i^2(1-\mu_i^n)\int F\psi_i$, which by (11) and (24), is seen to be equivalent to (28). Furthermore, from (8) and (21), we obtain $0 < 1/\lambda_i^2 < \int\int k^2(x,y)\,dxdy \leqq 2$, or $-1 < \mu_i < 1$. Thus, since the sequence $\{\lambda_i\}$ increases monotonically to infinity, it is seen that (29) is also satisfied. This completes the proof of Lemma 2.

LEMMA 3.

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} (c_{in} - c_i)^2 = \lim_{n\to\infty} \sum_{i=1}^{\infty} \lambda_i^2 (c_{in} - c_i)^2 = 0. \tag{30}$$

If a solution $g(x)$ *of* (1) *or* (2) *exists, then also*

$$\lim_{n\to\infty} \sum_{i=1}^{\infty} \lambda_i^4 (c_{in} - c_i)^2 = 0. \tag{31}$$

Proof. We first note that the series $\sum_{i=1}^{\infty} (c_{i0} - c_i)^2$ converges since we have, from Bessel's inequality $\sum_{i=1}^{\infty} (c_{i0} - c_i)^2 \leqq \int (F_0 - F)^2$; hence, $\sum_{i=1}^{\infty} (c_{in} - c_i)^2 = \sum_{i=1}^{\infty} \mu_i^{2n} (c_{i0} - c_i)^2$ is uniformly convergent in n, by (28) and the comparison test. Consequently, $\lim_{n\to\infty} \sum_{i=1}^{\infty} (c_{in} - c_i)^2 = \sum_{i=1}^{\infty} \lim_{n\to\infty} \mu_i^{2n} (c_{i0} - c_i)^2 = 0$. Similarly, applying Bessel's inequality to $f_0 - f$, and then to $g_0 - g$, when $g(x)$ is assumed to exist, we obtain (30) and (31), as desired.

LEMMA 4. *If the series* $\Gamma_0(x) = \sum_{i=1}^{\infty} w_i(x)$, *where the* $w_i(x)$ *are continuous functions, is absolutely and uniformly convergent, and if* $\Gamma_n(x) = \sum_{i=1}^{\infty} \mu_i^n w_i(x)$, $n = 0, 1, 2, \cdots$, *where* μ_i *satisfies condition* (29), *then the sequence* $\Gamma_n(x)$ *converges uniformly to* 0.

Proof. From the hypotheses on μ_i we have, for some sufficiently large r, $\mu_r \geqq |\mu_i|$, $r > i$. Also, considering the series for $\Gamma_0(x)$, given an $\epsilon > 0$, r can be chosen so large, and independent of x, that $\sum_{i=r+1}^{\infty} |w_i| < \epsilon/2$. Let r be chosen so that both conditions are satisfied. Further, we have $\sum_{i=1}^{r} |w_i| \leqq \sum_{i=1}^{\infty} |w_i| < M$, where M is an upper bound independent of x. Choose N sufficiently large so that $\mu_r^n < \epsilon/(2M)$ for $n > N$. Then $|\Gamma_n| \leqq \sum_{i=1}^{r} |\mu_i^n w_i| + \sum_{i=r+1}^{\infty} |\mu_i^n w_i|\, \mu_r^n M + \epsilon/2 < \epsilon$, when $n > N(\epsilon)$, as we wished to prove.

LEMMA 5. *If* $G_n(x)$ *can be expanded in a uniformly convergent series*

$$G_n(x) = \sum_{i=1}^{\infty} e_{in}\theta_i(x), \qquad n = 0, 1, 2, \cdots \tag{32}$$

in terms of the real, continuous, orthonormalized functions $\theta_i(x)$, $i = 1, 2, \cdots$ *and if* $G(x)$ *is piecewise-continuous, with* $e_i = \int G\theta_i$, *then necessary and sufficient conditions for the sequence* $G_n(x)$ *to converge in the mean to* $G(x)$ *are that* $\int G^2 dx = \sum_{i=1}^{\infty} e_i^2$ *and* $\lim_{n\to\infty} \sum_{i=1}^{\infty} (e_{in} - e_i)^2 = 0$.

Proof. Since the series (32) is uniformly convergent, we have $\int GG_n = \sum_{i=1}^{\infty} e_{in} \int G\theta_i = \sum_{i=1}^{\infty} e_{in}e_i$, and similarly $\int G_n{}^2 = \sum_{i=1}^{\infty} e_{in}{}^2$. Hence

$$\int (G_n - G)^2 = \int G^2 + \sum_{i=1}^{\infty} (e_{in} - e_i)^2 = \sum_{i=1}^{\infty} e_i^2. \tag{33}$$

Now suppose the conditions of the lemma to be satisfied. Then $\int (G_n - G)^2 = \sum_{i=1}^{\infty} (e_{in} - e_i)^2$ and consequently by hypothesis, $\lim \int (G_n - G)^2 = 0$. This proves the first part of the lemma.

Now suppose that $\lim \int (G_n - G)^2 = 0$. From (33), $\int G^2 dx \leqq \sum_{i=1}^{\infty} e_i^2 + \int (G_n - G)^2$ for all n. Hence $\int G^2 \leqq \sum_{i=1}^{\infty} e_i^2$. But, by Bessel's inequality, $\int G^2 \geqq \sum_{i=1}^{\infty} e_i^2$. Hence $\int G^2 = \sum_{i=1}^{\infty} e_i^2$. Then, from (33), $\sum_{i=1}^{\infty} (e_{in} - e_i)^2 = \int (G_n - G)^2$, whence $\lim \sum_{i=1}^{\infty} (e_{in} - e_i)^2 = 0$. This completes the proof.

We can now proceed to the proof of the convergence theorems.

Proof of Theorem 3. By the expansion theorem and (11) and (24), the series $F_n - F = \sum_{i=1}^{\infty} (c_{in} - c_i)\psi_i$, $n = 0, 1, 2, \cdots$ are absolutely and uniformly convergent in x. Hence, by Lemma 2, the series $\sum_{i=1}^{\infty} \mu_i{}^n (c_{i0} - c_i)\psi_i$ are also absolutely and uniformly convergent in x. Hence the conditions of Lemma 4 are satisfied and the sequence $\{F_n - F\}$ converges uniformly to zero; or by (23), $\{Kg_n\}$ converges uniformly to F, as we wished to prove.

Proof of Theorem 4. By Lemmas 1 and 3 all the conditions of Lemma 5 are satisfied by the functions $f_n(x)$ and $f(x)$. Hence by (23) the first part of the theorem, concerning the convergence in the mean of $\{kg_n\}$ to $f(x)$, is proved.

In the second part of the theorem, since $g(x)$ exists by hypothesis, the expansion theorem may be applied to $f(x)$ as well as to $f_n(x)$. Hence by (11) and Lemmas 1 and 2, the series $f_n - f = \sum_{i=1}^{\infty} \mu_i^n \lambda_i (c_{i0} - c_i) \phi_i(x)$, $n = 0, 1, 2, \cdots$ are absolutely and uniformly convergent in x, and the conditions of Lemma 4 are satisfied. Hence the sequence $\{f_n - f\}$ converges uniformly to zero, or, by (23), $\{kg_n\}$ converges uniformly to $f(x)$. This completes the proof.

Proof of Theorem 5. Since $g_0(x)$ is of the form (22), Lemmas 1 and 3 indicate that the conditions of Lemma 5 are satisfied by the functions $g_n(x)$ and $g(x)$. Hence the first part of the theorem, concerning convergence in the mean of $\{g_n\}$ to $g(x)$, is proved.

In the second part of the theorem, the expansion theorem is applicable to $g(x)$, by hypothesis. Hence, by (11) and Lemmas 1 and 2, the series $g_n - g = \sum_{i=1}^{\infty} \mu_i^n \lambda_i^2 (c_{i0} - c_i) \psi_i(x)$, $n = 0, 1, 2, \cdots$ are absolutely and uniformly convergent in x, and the conditions of Lemma 4 are satisfied. Hence the sequence $\{g_n\}$ converges uniformly to $g(x)$, as we wished to prove.

DAVID TAYLOR MODEL BASIN,
WASHINGTON, D. C.

REFERENCES.

[1] E. Hellinger and O. Toeplitz, "Integralgleichungen und Gleichungen mit unendlichvielen Unbekannten," *Encyklopädie der Mathematischen Wissenschaften* II C 13 (1928), pp. 1349.

[2] M. M. Munk, "Fluid Mechanics," Part II of *Aerodynamic Theory*, edited by W. F. Durand, vol. I, Julius Springer, Berlin (1935), Division C, pp. 269, equation (6. 1).

[3] E. Schmidt, "Zur Theorie der linearen und nichtlinearen Integralgleichungen," *Mathematische Annalen*, vol. 63 (1907), pp. 433-476.

[4] R. Courant und D. Hilbert, "Methoden der Mathematischen Physik," vol. 1, Kap. III, § 2.

[5] Reference [4], § 10.

[6] E. Picard, "Sur un théorème général relatif aux équations intégrales de première espèce et sur quelques problèmes de physique mathématique," *Rend. Circ. mat. Palermo*, vol. 29 (1910), pp. 79-97.

[7] A. Weinstein, "On axially symmetric flows," *Quarterly of Applied Mathematics*, vol. 5, No. 4 (1948), p. 433.

ON THE ASYMPTOTIC EVALUATION OF A CLASS OF MULTIPLE INTEGRALS INVOLVING A PARAMETER.*

By L. C. Hsu.

---

1. Introduction. The object of this paper is to determine the asymptotic behaviour of a class of multiple integrals of the form

$$I(\lambda) = \int_R \cdots \int \exp f(x_1, \cdots, x_n; \lambda)\, dx_1 \cdots dx_n,$$

as $\lambda \to \infty$. Here f is a real valued function of $x = (x_1, \cdots, x_n)$ defined in R and satisfying certain general conditions, R being a simply connected, finitely bounded, n-dimensional domain in the euclidean n-space.

We assume that, for every large λ, the function $f(x_1, \cdots, x_n; \lambda) = f(x; \lambda)$ attains an absolute maximum at an interior point $x = x(\lambda) = (x_1(\lambda), \cdots, x_n(\lambda))$ of R. The case in which $f = \lambda g(x_1, \cdots, x_n)$ and the function g takes a maximum at a boundary point of the closed domain R has been discussed in a previous paper [5]. But the method of [5] cannot be applied unaltered to the present case, because $x(\lambda)$ is now a variable point depending on the parameter λ. Thus in the present investigation we shall modify and extend Laplace-Haviland's method ([1], [2], [3]).

2. Notation and abbreviation. Throughout x, u, ξ, etc. denote respectively the points $(x_1, \cdots, x_n)$, $(u_1, \cdots, u_n)$, $(\xi_1, \cdots, \xi_n)$, etc. in the euclidean n-space, f_k denotes the partial derivative of $f(x; \lambda)$ with respect to x_k, and f_{ik} the second order partial derivative of f with respect to x_k and x_i. Moreover $H_k[-f]$ denotes the Hessian of $-f(x; \lambda)$ with regard to the first k variables $x_1, \cdots, x_k$, viz., $H_k[-f(x; \lambda)] = \det \| -f_{\alpha\beta}(x; \lambda) \|$, where $1 \leqq \alpha, \beta \leqq k$ $(k = 1, \cdots, n)$.

By $(x; \lambda) \to (\xi; \infty)$ we mean that x, λ tend to ξ, ∞ respectively and independently. Thus, for instance, $f(x; \lambda) = o(g(x; \lambda))$ as $(x; \lambda) \to (\xi; \infty)$ means that for any given $\epsilon > 0$, there exist a small number $\delta = \delta(\epsilon) > 0$ and a large number $N = N_{(\epsilon)} > 0$ such that $| f/g | < \epsilon$ fhenever $| x_k - \xi_k | < \delta$ $(k = 1, \cdots, n)$ and $\lambda > N$. In an analogous manner we may define $f = O(g)$, $f \sim g$ as $(x; \lambda) \to (\xi; \infty)$.

As a convention we assume that any statement involving index i or k

---

* Received June 8, 1950.

is true for all $i, k = 1, \cdots, n$. Thus, for example, $f_k = 0$ represents not a single equation but a system of simultaneous equations $f_k(x;\lambda) = 0$ $(k = 1, \cdots, n)$.

3. Statement of theorems. For convenience we shall say that $f(x;\lambda)$ belongs to the class C^2 with parameter λ if f, f_k, f_{ik} are continuous functions with respect to the variables x and λ.

THEOREM 1. *Let* $f(x;\lambda) \varepsilon C^2$ *for* $x \varepsilon R$ *with parameter* λ *such that*

(i) *for every large* λ *the function* $f(x;\lambda)$ *attains an absolute maximum at an interior point* $x(\lambda) = (x_1(\lambda), \cdots, x_n(\lambda))$ *of* R, *so that* $f_k = 0$, $H_k[-f] > 0$ *at* $x = x(\lambda)$, *where* $x(\lambda)$ *tends to an interior point* ξ *of* R *as* $\lambda \to \infty$;

(ii) *for* λ *large* $H_k[-f] > 0$ *hold throughout* R;

(iii) *for* $(x;\lambda) \to (\xi;\infty)$ *we have* $f_{ik}(x;\lambda) \sim f_{ik}(\xi;\lambda)$, $H_k[-f] \to \infty$ *and* $\log H_n[-f] = o(H_k[-f]/H_{k-1}[-f])$, *where* $H_0[-f] = 1$.

Then for $\lambda \to \infty$ *we have*

$$(1)\qquad \int_R \cdots \int \exp f(x;\lambda)\, dx_1 \cdots dx_n \sim \exp[f(x_1(\lambda), \cdots, x_n(\lambda);\lambda)]((2\pi)^n/H_n[-f(\xi;\lambda)])^{1/2}.$$

Note that the right-hand side of (1) does not involve R. As a matter of fact, our proof of Theorem 1 depends essentially on an asymptotic investigation of the integral in a small neighbourhood of $x = x(\lambda)$ or of $x = \xi$. Thus obviously the domain of integration R may be replaced by a larger domain D, provided that

(iv) *for* λ *large we have* l. u. b. $f(y_1, \cdots, y_n;\lambda) \leqq$ g. l. b. $f(x_1, \cdots, x_n;\lambda)$ *in which* $x \varepsilon R$, $y = (y_1, \cdots, y_n) \varepsilon D - R$, *where* R *may be an arbitrary small domain containing* ξ *but fixed.*

Moreover by an application of the mean value theorem for the integral calculus we easily establish the following

THEOREM 2. *If* $\phi(x)$ *is absolutely integrable over a finite domain* D, *if* $f(x;\lambda)$ *satisfies all the hypotheses of Theorem* 1 *together with* (iv) *in which* R *is a fixed neighbourhood of* ξ, *and if* $\phi(x)$ *is continuous at* $x = \xi$ *with* $\phi(\xi) \neq 0$, *then we have*

$$(2)\qquad \int_D \cdots \int \phi(x) \exp f(x;\lambda)\, dx_1 \cdots dx_n \sim \exp[f(x(\lambda);\lambda)]\phi(\xi)((2\pi)^n/H_n[-f(\xi;\lambda)])^{1/2}.$$

Evidently when $n = 1$ condition (iii) reduces to the following

$$(\text{iii})^1 \qquad f''_{xx}(x;\lambda) \sim f''_{xx}(\xi;\lambda), \qquad -f''_{xx}(x;\lambda) \to \infty \text{ as } (x;\lambda) \to (\xi;\infty).$$

This condition is automatically satisfied for the simplest case $f(x;\lambda) = \lambda h(x)$ with $h'(\xi) = 0$, $h''(\xi) < 0$. Thus it is clear that the 1-dimensional case of Theorem 2 is actually a direct generalization of the classical Laplace theorem for an asymptotic integral (see [1; 277-280], [3; 78]). Similarly it is also easily seen that Lemma 1 of [5; 624] is implied by our Theorem 1 by taking $f(x_1, \cdots, x_n;\lambda) = \lambda g(x_1, \cdots, x_n)$.

4. **Proof of theorems.** The following Lemma 1 will play an important rôle in our proof of Theorem 1:

LEMMA 1. *Let* $A = [a_{ik}]$ *be a symmetrical matrix of the definite form* $xAx' = \Sigma\Sigma a_{ik}x_i x_k$ *in* n *variables. Then*

$$(3) \quad xAx' = \sum_{k=1}^{n} (A_{k-1}{}^{(k-1)} \cdot A_k{}^{(k)})^{-1} \; (A_k{}^{(k)} x_k + A_{k+1}{}^{(k)} x_{k+1} + \cdots + A_n{}^{(k)} \cdot x_n)^2,$$

where x' *denotes the transposed of* x, *and* $A_0{}^{(0)} = 1$, $A_s{}^{(1)} = a_{s1}$,

$$(4) \qquad A_s{}^{(k)} = \begin{vmatrix} a_{11} & \cdots & a_{1k} \\ \cdot & \cdots & \cdot \\ a_{k-1\,1} & \cdots & a_{k-1\,k} \\ a_{s1} & \cdots & a_{sk} \end{vmatrix}, \qquad (1 \leq k \leq s \leq n).$$

The explicit formula (3) is actually a direct consequence of Darboux's reduction of quadratic forms [4; 191-192]. It may also be proved by repeated application of Lagrange's reduction and by induction. The detailed proof will be omitted here.

In what follows we shall say that a hypersurface defined by $z = f(x_1, \cdots, x_n)$ for $x \,\varepsilon\, R$ is concave downward in R, if the second order directional derivative of f is negative definite throughout R.

LEMMA 2. *Let* $f(x_1, \cdots, x_n;\lambda)$ *satisfy hypotheses* (i) *and* (ii) *of Theorem* 1. *Then for every fixed large* λ *the hypersurface represented by* $z = f(x;\lambda)$ *is concave downward in* R *and attains its highest point at* $x = x(\lambda)$.

Let us take an arbitrary line of the euclidean n-space, say $\Gamma_{(\theta_1,\ldots,\theta_n)}$, passing through a point $(x_1{}^0, \cdots, x_n{}^0)$ of R and having directional cosines

$\cos\theta_1, \cdots, \cos\theta_n$, where θ_k is the angle made by Γ and the x_k-axis, so that $\Sigma\cos^2\theta_k = 1$ and the equation of Γ may be written as

$$(x_1 - x_1^0)/\cos\theta_1 = (x_2 - x_2^0)/\cos\theta_2 = \cdots = (x_n - x_n^0)/\cos\theta_n.$$

It is immediately found that the first and second directional derivatives of $z = f(x_1, \cdots, x_n; \lambda)$ with respect to the direction Γ_θ are as follows

$$(5)\qquad \partial f/\partial\Gamma_\theta = \sum_{k=1}^{n} f_k \cos\theta_k, \qquad \partial^2 f/\partial\Gamma_\theta^2 = \sum_{i=1}^{n}\sum_{k=1}^{n} f_{ik}\cos\theta_i\cos\theta_k.$$

Obviously we may express $\partial^2 f/\partial\Gamma_\theta^2 = yHy'$, where $H = [f_{ik}]$, $y = (y_1, \cdots, y_n) = (\cos\theta_1, \cdots, \cos\theta_n)$, and $y' = \{y_1, \cdots, y_n\}$ is the transposed of y. Hence applying Lemma 1 with $a_{ik} = f_{ik}$ and using hypothesis (ii) we get

$$(6)\qquad \partial^2 f/\partial\Gamma_\theta^2 = yHy' < 0,$$

for every large λ, say $\lambda > M$. Our lemma thus follows from (i) and (6).

Proof of Theorem 1. Let us now keep $\lambda(> M)$ fixed and denote

$$\int_R \cdots \int \exp f(x;\lambda)\,dR = \int_U \cdots \int \exp f(x;\lambda)dx + \int_{R-U} \cdots \int \exp f(x;\lambda)dx = I_1(\lambda) + I_2(\lambda),$$

where U is a neighbourhood of $x(\lambda) = (x_1(\lambda), \cdots, x_n(\lambda))$ whose volume depends on λ, $R - U$ denotes the complementary region of U with respect to the whole domain R. For convenience, we take $U = U_{(\epsilon)}$ to be a hypercube having the variable length $2\epsilon(\lambda)$ on each side and containing $x(\lambda)$ as its centre. By Taylor's formula with remainder we have

$$(7)\qquad \begin{aligned} J(\lambda) &= I_1(\lambda)\exp[-f(x_1(\lambda), \cdots, x_n(\lambda); \lambda)] \\ &= \int_U \cdots \int \exp[\tfrac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n} f_{ik}(X_1, \cdots, X_n; \lambda)(x_i - x_i(\lambda))(x_k - x_k(\lambda))] \\ &\qquad \times dx_1 \cdots dx_n, \end{aligned}$$

where

$$(8)\qquad |X_k - x_k(\lambda)| < |x_k - x_k(\lambda)| \leq \epsilon(\lambda).$$

Since all X_k depend on x we may write $X_k = X_k(x)$. Moreover for fixed x we denote

$$(9)\qquad x_k - x_k(\lambda) = u_k, \qquad -f_{ik}(X_1, \cdots, X_n; \lambda) = b_{ik}, \qquad [b_{ik}] = B.$$

Then the integral $J(\lambda)$ may be written as

$$J(\lambda) = \int_{-\epsilon(\lambda)}^{\epsilon(\lambda)} \cdots \int_{-\epsilon(\lambda)}^{\epsilon(\lambda)} \exp(-\tfrac{1}{2}uBu')\, du_1 \cdots du_n, \tag{10}$$

where $u = (u_1, \cdots, u_n)$ and $u' = \{u_1, \cdots, u_n\}$.

For asymptotic evaluation of $J(\lambda)$ we need the following

LEMMA 3. *Let* $-f_{ik}(x(\lambda); \lambda) = a_{ik}$ *and* $[a_{ik}] = A$. *Then*

$$\lim_{u_k \to 0} (uBu')/(uAu') = 1. \tag{11}$$

From condition (ii) it is clear that uAu' is a positive definite form in n variables for $\lambda > M$. On the other hand, we see that the functional elements of B are continuous at $x = x(\lambda)$, i. e. $b_{ik} = -f_{ik}(X_1(x), \cdots, X_n(x); \lambda) \to -f_{ik}(x_1(\lambda), \cdots, x_n(\lambda); \lambda) = a_{ik}$ as $x \to x(\lambda)$ or correspondingly $u \to (0, \cdots, 0)$. Thus when u tends to $(0, \cdots, 0)$ through any path, say $u_k = \rho_k t$ as $t \to 0$, ρ_k being constants not all zero, we always have $\lim (uBu'/uAu') = \lim (\rho B\rho'/\rho A\rho') = 1$ where $\rho = (\rho_1, \cdots, \rho_n)$. Our lemma is therefore proved.

Now suppose that we choose ϵ in such a way that $\epsilon(\lambda) \to 0$ as $\lambda \to \infty$. Then from (8) and (9) it follows that $u = (u_1, \cdots, u_n) \to (0, \cdots, 0)$ as $\lambda \to \infty$, so that by Lemma 3 we may write

$$uBu' = uAu'(1 + \delta(u)), \tag{12}$$

where $\delta(u) \to 0$ as $\lambda \to \infty$. It is known that uAu' can be reduced to the sum of n squares by means of a congruent linear transformation $u' = Tv'$ with $T'AT = I$ so that

$$uAu' = vT'ATv' = vv' = \sum_{k=1}^{n} v_k^2, \tag{13}$$

where $v = (v_1, \cdots, v_n)$ and the Jacobian of the transformation is given by $|\partial(u)/\partial(v)| = |T| = |A|^{-1/2}$.

Under the transformation T the neighbourhood U of w is transformed to a neighbourhood V of v. Set

$$M(\epsilon) = \max_{|u_k| \leq \epsilon} \{\delta(u)\}, \qquad m(\epsilon) = \min_{|u_k| \leq \epsilon} \{\delta(u)\}. \tag{14}$$

Clearly from (10), (12), (13) and (14) we have

$$\begin{aligned} J(\lambda) &\leq |A|^{-1/2} \int_V \cdots \int \exp[(-\tfrac{1}{2}vv')(1 + m(\epsilon))]\, dv_1 \cdots dv_n = J(\lambda), \\ J(\lambda) &\geq |A|^{-1/2} \int_V \cdots \int \exp[(-\tfrac{1}{2}vv')(1 + M(\epsilon))]\, dv_1 \cdots dv_n = J(\lambda), \end{aligned} \tag{15}$$

where v_k^2 can be written explicitly by means of Lemma 1, viz.

$$v_k^2 = (A_{k-1}^{(k-1)} \cdot A_k^{(k)})^{-1}(A_k^{(k)}u_k + \cdots + A_n^{(k)} \cdot u_n)^2. \tag{16}$$

What we shall show next is that there is a positive valued function $\epsilon(\lambda)$ satisfying the following conditions:

I) $\lim \epsilon(\lambda) = 0$ as $\lambda \to \infty$;

II) the integration limits (upper and lower limits) for each v_k of $\bar{J}$ or $\underline{J}$ tend to $\pm\infty$ as $\lambda \to \infty$;

III) $\lim (I_2(\lambda)/I_1(\lambda)) = 0$ as $\lambda \to \infty$.

We are now going to construct such a suitable $\epsilon(\lambda)$-function. For every fixed $\lambda > M$ and fixed $x \varepsilon R$, let us denote

$$\mu(x;\lambda) = \min \left\{ \frac{H_1[-f]}{H_0[-f]}, \frac{H_2[-f]}{H_1[-f]}, \cdots, \frac{H_n[-f]}{H_{n-1}[-f]} \right\}, \tag{17}$$

where $H_0[-f] = 1$. Then our $\epsilon(\lambda)$-function may simply be defined as follows

$$\epsilon(\lambda) = K_0 \cdot \left(\frac{\log H_n[-f(x_1(\lambda), \cdots, x_n(\lambda); \lambda)]}{\mu(x_1(\lambda), \cdots, x_n(\lambda); \lambda)} \right)^{1/2}, \tag{18}$$

where K_0 is a suitable positive number.

It is clear that the $\epsilon(\lambda)$-function just defined satisfies condition I) by our hypothesis (iii) that $\log H_n[-f] = o(H_k[-f]/H_{k-1}[-f])$ as $(x;\lambda) \to (\xi; \infty)$. For verification of II) one needs only to substitute $\pm\epsilon(\lambda)$ for u_k's on the right-hand side of (16), e.g. if $u_k = \pm\epsilon(\lambda)$, $u_{k+1} = \cdots = u_n = 0$, then by (17), (18) and hypothesis (iii) we have

$$\begin{aligned} v_k^2 &= (A_{k-1}^{(k-1)} \cdot A_k^{(k)})^{-1}(A_k^{(k)}\epsilon(\lambda))^2 \\ &= \epsilon(\lambda)^2 H_k[-f(x(\lambda);\lambda)]/H_{k-1}[-f(x(\lambda);\lambda)] \\ &\geq K_i\epsilon(\lambda)^2\mu(x(\lambda);\lambda) = K_0^2 \cdot K_1 \log H_n[-f(x(\lambda);\lambda)] \to \infty, \end{aligned} \tag{19}$$

as $\lambda \to \infty$, where K_1 is a certain positive number (≥ 1) implied by (17). This shows that the range of $|v_k|$ tends to infinity with λ. Note that our previous linear transformation T is non-singular and leaves the convexity of the domain $U_{(\epsilon)}$ invariant. It is clear that the corresponding domain V ($= V_{\epsilon(\lambda)}$) of v will have no finite boundary or definite bounds as $\lambda \to \infty$, i.e. the integration limits for each v_k will become $\pm\infty$ as $\lambda \to \infty$. Thus by noticing $m(\epsilon) \to 0$, $M(\epsilon) \to 0$,as $\lambda \to \infty$) and making use of the complete probability integral, we easily obtain (cf. [5; 626])

$$\lim_{\lambda\to\infty} |A|^{\frac{1}{2}} J(\lambda) = \lim_{\lambda\to\infty} |A|^{\frac{1}{2}} J(\lambda) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp(-\tfrac{1}{2}vv')\,dv_1 \cdots dv_n = (2\pi)^{\frac{1}{2}n}.$$

From (15), (7) and by hypothesis (iii) (i. e. $f_{ik}(x;\lambda) \sim f_{ik}(\xi;\lambda)$ as $(x;\lambda) \to (\xi;\infty)$) we may therefore infer that

$$\begin{aligned} I_1(\lambda) &\sim \exp[f(x_1(\lambda),\cdots,x_n(\lambda);\lambda)]\,|\,A\,|^{-\frac{1}{2}}(2\pi)^{\frac{1}{2}n} \\ &\sim \exp[f(x(\lambda);\lambda]\,\{(2\pi)^n/H_n[-f(\xi;\lambda)]\}^{\frac{1}{2}}. \end{aligned} \tag{20}$$

Finally we have to verify condition III). Recall that the hypersurface $z = f(x;\lambda)$ (for fixed $\lambda > M$) defined on R is concave downward (by Lemma 2) and that U is a neighbourhood of $x(\lambda)$. Hence it is clear that in the complementary region $R - U$ the function $f(x;\lambda)$ can reach its absolute maximum only on the boundary of U. Denote by $\tilde{U}$ the boundary of U and by $\| R \|$ the volume of R. Let $h(\lambda)$ denote the maximum or the least upper bound of $\exp f(x;\lambda)$ as x varies over $\tilde{U}$. Obviously we have

$$I_2(\lambda) \leqq \| R \| \cdot h(\lambda). \tag{21}$$

Since U is a hypercube of side-length $2\epsilon(\lambda)$ containing the point $x(\lambda) = (x_1(\lambda),\cdots,x_n(\lambda))$ as its centre, we see that every point of $\tilde{U}$ may be written in the form

$$x^0 = (x_1(\lambda) + \gamma_1\cdot\epsilon(\lambda),\cdots,x_n(\lambda) + \gamma_n\cdot\epsilon(\lambda)), \tag{22}$$

where $|\gamma_i| \leq 1$ $(i = 1,\cdots,n)$ and $|\gamma_k| = 1$ for at least one k $(1 \leq k \leq n)$. Now suppose that x^0 is a point at which $\exp f(x;\lambda)$ takes the maximum $h(\lambda)$. We may write (cf. (7) and (8)).

$$h(\lambda) = \exp[f(x(\lambda);\lambda) + \tfrac{1}{2}\epsilon^2 \sum_{i=1}^{n}\sum_{k=1}^{n} f_{ik}(X_1,\cdots,X_n;\lambda)\gamma_i\gamma_k], \tag{23}$$

where $|X_k - x_k(\lambda)| < \gamma_k\cdot\epsilon(\lambda) \to 0$ with ϵ. Hence using Lemmas 3 and 1, we have, for λ large (cf. (16)),

$$\begin{aligned} -\sum_{i=1}^{n}\sum_{k=1}^{n} f_{ik}(X_1,\cdots,X_n;\lambda)\gamma_i\gamma_k &\sim \sum_{i=1}^{n}\sum_{k=1}^{n}(-)f_{ik}(x(\lambda);\lambda)\gamma_i\gamma_k \\ &= \gamma A\gamma' = \sum_{k=1}^{n}(A_{k-1}{}^{(k-1)}\cdot A_k{}^{(k)})^{-1}(A_k{}^{(k)}\gamma_k + \cdots + A_n{}^{(k)}\cdot\gamma_n)^2 > 0. \end{aligned} \tag{24}$$

Note that γ_k's are constants not all zero. If $\gamma_n \neq 0$, then the positive definite form $\gamma A\gamma'$ contains the term $(A_n{}^{(n)}/A_{n-1}{}^{(n-1)})\cdot\gamma_n^2 = (H_n[-f]/H_{n-1}[-f])\cdot\gamma_n^2$. Similarly if $\gamma_{k+1} = \cdots = \gamma_n = 0$ but $\gamma_k \neq 0$, then $\gamma A\gamma'$ contains the positive term $(H_k[-f]/H_{k-1}[-f])\cdot\gamma_k^2$. Hence from (24) and (17) we see that in any case we have a positive constant, say $K_2 > 0$, such that

$$(25) \qquad -\sum_{i=1}^{n} \sum_{k=1}^{n} f_{ik}(X_1, \cdots, X_n; \lambda)\gamma_i\gamma_k > K_2 \cdot \mu(x(\lambda);\lambda),$$

when λ is large. Thus by (20), (21), (23), (25) and (18) we have

(26)

$$\begin{aligned} 0 &< I_2(\lambda)/I_1(\lambda) \\ &< K_3 \| R \| \cdot A |^{\frac{1}{2}}/(2\pi)^{n/2} \exp\left[\tfrac{1}{2}\epsilon(\lambda)^2 \sum_{i=1}^{n} \sum_{k=1}^{n} f_{ik}(X_1, \cdots, X_n; \lambda)\gamma_i\gamma_k\right] \\ &< K_4 | A |^{\frac{1}{2}} \exp[-\tfrac{1}{2}K_2 \cdot \epsilon(\lambda)^2 \cdot \mu(x(\lambda);\lambda)] \\ &= K_4 | A |^{\frac{1}{2}} \exp\{-\tfrac{1}{2} K_2 \cdot K_0^2 \log H_n[-f(x(\lambda);\lambda)]\} \\ &= K_4 | A |^{\frac{1}{2}} \{H_n[-f(x(\lambda);\lambda)]\}^{-\frac{1}{2}K_0^2 \cdot K_2} \\ &= K_4 \{H_n[-f(x(\lambda);\lambda)]\}^{\frac{1}{2}(1-K_0^2 \cdot K_2)}, \end{aligned}$$

where K_3, K_4 are certain positive numbers. Since the positive constant K_0 of (18) can be chosen freely, we may assume that $K_0^2 K_2 > 1$. Hence the expression (26) leads at once to the conclusion that $I_2(\lambda)/I_1(\lambda) \to 0$ as $\lambda \to \infty$. Our Theorem 1 is therefore completely proved by (20) and III).

Proof of Theorem 2. Evidently the domain of integration R in the formula (1) can be replaced by D, if hypothesis (iv) is added. For we need only replace $\| R \|$ by $\| D \|$ in (21). Now let W denote an arbitrary small region (neighbourhood) of $x = \xi$. Let $J_1(\lambda)$, $J_2(\lambda)$ denote n-tiple integrals of $[\phi(x) - \phi(\xi)] \exp f(x;\lambda)$ taken over the domains W and $D - W$ respectively. We may express

$$\int_D \cdots \int \phi(x) \exp f(x;\lambda)\, dx_1 \cdots dx_n = \phi(\xi) I(\lambda) + J_1(\lambda) + J_2(\lambda),$$

where $I(\lambda)$ is defined as in § 1 (save that now the domain R is replaced by D). Since $\phi(x)$ is continuous at $x = \xi$ the difference $| \phi(x) - \phi(\xi) |$ may be made as small as we like, provided that W is sufficiently small. Thus the First Mean Value Theorem for integrals shows at once that $J_1(\lambda) = o(I(\lambda))$. Denote by $h(\lambda)$ the maximum of $\exp f(x;\lambda)$ as x varies over the boundary of W. Clearly we have

$$J_2(\lambda) \leq h(\lambda) \int_{D-W} \cdots \int | \phi(x) - \phi(\xi) |\, dx_1 \cdots dx_n,$$

which now replaces (21). Thus the same calculation as that of (26) shows that $J_2(\lambda) = o(I_1(\lambda))$, $J_2(\lambda) = o(\phi(\xi) I(\lambda))$ and this completes the proof of Theorem 2.

5. Simple consequences and discussion. It is clear that in applying the asymptotic formulas (1) and (2) we need first determine the functional solution $(x_1(\lambda), \cdots, x_n(\lambda))$ of the simultaneous equations $f_k(x_1, \cdots, x_n; \lambda) = 0$ for large λ. In case we require only the approximate values of $x_k(\lambda)$'s, we may obtain them directly by solving $f_k(x_1, \cdots, x_n; \lambda) = 0$ for some fixed large λ. Generally speaking, it is not necessarily true that $(f(x(\lambda); \lambda) - f(\xi; \lambda)) \to 0$ as $\lambda \to \infty$, so that we must not replace $\exp f(x(\lambda); \lambda)$ by $\exp f(\xi; \lambda)$ in our formulas (1) and (2), unless some further condition is imposed.

For $n = 2$ our Theorems 1 and 2 reduce to the case of double integrals. In such a case the conditions (ii) and (iii) are equivalent to the following

(ii)^1 $\Delta_1 = -f_{11}(x_1, x_2; \lambda) > 0$, $\Delta_2 = f_{11}f_{22} - f_{12}^2 > 0$ *hold in* R, *where* $\Delta_1, \Delta_2 \to \infty$ *as* $(x_1, x_2) \to (\xi_1, \xi_2)$ *and* $\lambda \to \infty$;

(iii)^1 $\log \Delta_2 = o(\Delta_1)$, $\log \Delta_2 = o(\Delta_2/\Delta_1)$, $f_{ik}(x_1, x_2; \lambda) \sim f_{ik}(\xi_1, \xi_2; \lambda)$ *as* $(x_1, x_2) \to (\xi_1, \xi_2)$ *and* $\lambda \to \infty$.

And the formula (1) becomes

$$(1)^1 \quad \int_R\!\!\int \exp f(x_1, x_2; \lambda) dR \sim \frac{(2\pi) \exp f(x_1(\lambda), x_2(\lambda); \lambda)}{\{f_{11}(\xi_1, \xi_2; \lambda) f_{22}(\xi_1, \xi_2; \lambda) - f_{12}(\xi_1, \xi_2; \lambda)^2\}^{1/2}}.$$

As an application of Theorem 1, we now consider multiple integrals of the form

$$G(\lambda) = \int_R \cdots \int \exp[\lambda g(x_1, \cdots, x_n; \lambda)] dx_1 \cdots dx_n,$$

where $g(x; \lambda)$ is a bounded function for λ large and belongs to C^2 for $x \in R$.

If furthermore $g(x; \lambda)$ satisfies the following conditions:

(a) *for every large* λ *the function* $g(x; \lambda)$ *assumes an absolute maximum at an interior point* $x(\lambda) = (x_1(\lambda), \cdots, x_n(\lambda))$ *of* R, *where* $(x_1(\lambda), \cdots, x_n(\lambda))$ *tends to an interior point* ξ *as* $\lambda \to \infty$;

(b) $H_k[-g(x; \lambda)] > c > 0$ *hold in* R *for* λ *large*;

(c) $g_{ik}(x; \lambda) \sim g_{ik}(\xi; \lambda)$ *as* $(x; \lambda) \to (\xi; \infty)$,

then obviously Theorem 1 is applicable and we have

$$(1)'' \qquad G(\lambda) \sim \frac{(\exp g(x_1(\lambda), \cdots, x_n(\lambda); \lambda))^\lambda}{\{H_n[-g(\xi_1, \cdots, \xi_n; \lambda)]\}^{1/2}} \left(\frac{2\pi}{\lambda}\right)^{n/2}.$$

For it is quite clear that in our present case the conditions $H_k[-f] \to \infty$ and $\log H_n[-f] = o(H_k[-f]/H_{k-1}[-f])$ are automatically satisfied with $f = \lambda g(x; \lambda)$. Concrete examples are easily found for illustrating the use of formula (1)$''$.

I wish to express my hearty thanks to the referee for valuable comments and suggestions. Finally I have to mention that the analytical method adopted here is almost the same as that already developed in my paper [6], but the results presented in this paper seem much more general than those contained in [6].

KING'S COLLEGE, ABERDEEN, SCOTLAND.

REFERENCES.

[1] D. V. Widder, *The Laplace Transform* (Princeton University Press, 1946), Chap. 7, pp. 277-280.

[2] E. K. Haviland, "On an asymptotic expression for a certain integral," *Quarterly Journal of Mathematics* (Oxford), vol. 7 (1936), pp. 152-157.

[3] G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis* (New York, 1945), Kap. 5, pp. 77-82.

[4] R. F. Scott and G. B. Mathews, *Theory of Determinants and Their Applications* (Cambridge University Press, 1904).

[5] L. C. Hsu, A Theorem, "On the asymptotic behavior of a multiple integral," *Duke Mathematical Journal*, vol. 15 (1948), pp. 623-632.

[6] ———, "The asymptotic behavior of a kind of multiple integrals," to appear in the *Quarterly Journal of Mathematics* (Oxford).

THE NUMBER OF L^2-SOLUTIONS OF $x'' + q(t)x = 0$.*

By Philip Hartman.[1]

1. Let $q = q(t)$ be a real-valued continuous function for $0 \leqq t < \infty$. The differential equation

$$x'' + qx = 0 \tag{1}$$

and a homogeneous boundary condition at $t = 0$ determine a self-adjoint boundary value problem if and only if (1) is of the *Grenzpunkt* type, that is, if and only if the number of linearly independent solutions of class $L^2 = L^2(0, \infty)$ is less than 2; cf. [8]. This paper is principally concerned with conditions on q which are sufficient to assure that (1) is of *Grenzpunkt* type.

Let $N(t)$ denote the number of zeros on the interval $(0, t)$ of a non-trivial solution of (1); so that, up to an additive correction -1, 0 or 1, the integer $N(t)$ is independent of the choice of the particular solution of (1) determining $N(t)$. It was shown in [4] that a sufficient condition for (1) to be of *Grenzpunkt* type is that

$$\liminf_{t \to \infty} N(t)/t^2 < \infty. \tag{2}$$

It was also shown that, if $q^+(t) = \max(q(t), 0)$, then

$$N(t) \leqq (t \int_0^t q^+(s)\,ds)^{\frac{1}{2}} + O(1), \tag{3}$$

as $t \to \infty$. Thus, (2) holds if

$$\int^t q^+(s)\,ds = O(t^3), \tag{4}$$

as $t \to \infty$, and so, in particular, if

$$q(t) \leqq Ct^2 \tag{5}$$

holds for large values of t and some constant C.

---

* Received October 14, 1950.

[1] John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

The sufficiency of (5) was also obtained in [7]. The sufficient condition (5) was generalized to the conditions

$$(6) \qquad q(t) \leqq Q^2(t)$$

and

$$(7) \qquad \int^{\infty} Q^{-1}(s)\,ds = \infty,$$

where $Q(t) > 0$ is a function subject either to the condition that

(8) $Q^{-1}(t)$ satisfies a uniform Lipschitz condition

for large t, say, const. $\leqq t < \infty$ (cf. [5,] [6]) or to the condition that

(9) $Q(t)$ is monotone for large t

(cf. [6]). [In [5], a differential equation more general than (1) is considered; the corresponding assumptions reduce to (6), (7), (8) and (9), but the hypothesis (9) is not used in the proof.] The methods in [5], [6] are adaptations of those first used in [9] for the case $Q(t) = \text{const.}$ and then in [3] for the cases where the sign of equality holds in (6).

The generalizations of (5) to (6), (7) do not imply the sufficiency of (2), or even that of (4). It would, therefore, be of interest to obtain sufficient criteria involving $N(t)$, which would include those involving (2), hence (4), and (6), (7).

2. In this direction, a first guess might be that the t, whose square occurs in (2), can be replaced by a smooth monotone $Q(t)$ satisfying (7) (at least when lim inf is replaced by lim sup). However, this is not the case as is shown by the example $q(t) = t^2 \log^{2+4\epsilon} t$, where $0 < \epsilon \leqq \frac{1}{2}$. It is easily seen that the corresponding differential equation (1) is not of *Grenzpunkt* type. For, by known asymptotic formulae (cf. [10]), every solution $x(t)$ satisfies the estimate $x(t) = O(t^{-\frac{1}{2}} \log^{-\frac{1}{2}-\epsilon} t)$ and is therefore of class L^2, since $\epsilon > 0$. But $\pi N(t) \sim \int^t q^{\frac{1}{2}}(s)\,ds \sim \text{const.}\, t^2 \log^{1+2\epsilon} t$, cf. [2]; hence,

$$\limsup_{t\to\infty} N(t)/Q^2(t) < \infty, \text{ where } Q(t) = t \log^{\frac{1}{2}+\epsilon} t$$

satisfies (7) if $\epsilon \leqq \frac{1}{2}$.

3. The general criteria for (1) to be of *Grenzpunkt* type, to be proved below, are contained in the following assertion (and its proof):

(I) *The differential equation* (1) *is of Grenzpunkt type if there exists a positive continuous function* $Q(t)$ *defined for large positive* t, *of bounded variation* (*on finite intervals*) *and satisfying*

$$(10)\qquad \limsup_{t\to\infty}\,\{2\int^{t} Q^{-1}(s)\,ds-\int^{t} Q^{-2}(s)\,dN(s)-Q^{-2}(t)-\int^{t} |\,dQ^{2}(s)\,|\}=\infty.$$

The proof of (I) starts with a device employed in [4]. Let $x=x(t)$ and $x=y(t)$ be two linearly independent solutions of (1) normalized by the Wronskian condition $y'x-yx'=1$ and let $\theta(t)$ be the continuous function defined by

$$(11)\qquad \theta(t)=\arctan y(t)/x(t),\qquad\qquad |\,\theta(0)\,|\leqq\pi.$$

Then $\theta'(t)=(x^2(t)+y^2(t))^{-1}>0$. The differential equation (1) is of *Grenzpunkt* type if and only if at least one of the functions $x(t)$, $y(t)$ is not of class L^2, that is, if and only if

$$(12)\qquad \int^{\infty} dt/\theta'(t)=\infty.$$

The inequality $1=pp^{-1}\leqq\frac{1}{2}p^2+\frac{1}{2}p^{-2}$, when applied to $p=(\theta'(t)/Q(t))^{\frac{1}{2}}$, gives

$$(13)\qquad 1/\theta'\geqq 2/Q-\theta'/Q^2;$$

so that

$$(14)\qquad \int_{a}^{t} ds/\theta'(s)\geqq 2\int_{a}^{t} Q^{-1}(s)\,ds-\int_{a}^{t} Q^{-2}(s)\theta'(s)\,ds,$$

if a is sufficiently large and $a<t$. The last integral can be integrated by parts:

$$(15)\qquad -\int_{a}^{t} Q^{-2}(s)\theta'(s)\,ds=-\theta(t)Q^{-2}(t)+\theta(a)Q^{-2}(a)+\int_{a}^{t}\theta(s)\,dQ^{-2}(s).$$

Since the zeros of $x(t)$, $y(t)$ separate each other, it is clear that

$$(16)\qquad |\,\theta(t)-\pi N(t)\,|\leqq\pi.$$

Hence the right hand side of (15) is minorized by π times

$$(17)\qquad -N(t)Q^{-2}(t)+N(\alpha)Q^{-2}(\alpha) + \int_{\alpha}^{t} N(s)\,dQ^{-2}(s) - Q^{-2}(t) - Q^{-2}(\alpha) - \int_{\alpha}^{t} |\,dQ^{-2}(s)\,|,$$

which can be contracted into

$$(18)\qquad -\int_{\alpha}^{t} Q^{-2}(s)\,dN(s) - Q^{-2}(t) - Q^{-2}(\alpha) - \int_{\alpha}^{t} |\,dQ^{-2}(s)\,|.$$

Consequently,

$$\pi^{-1}\int_{\alpha}^{t} ds/\theta'(s) \geqq 2\pi^{-1}\int_{\alpha}^{t} Q^{-1}(s)\,ds - \int_{\alpha}^{t} Q^{-2}(s)\,dN(s) - Q^{-2}(t) - Q^{-2}(\alpha) - \int_{\alpha}^{t} |\,dQ^{-2}(s)\,|.$$

(If $Q(s)$ is replaced by $\pi Q(s)$ and if the last inequality is multiplied by π^{-2}, the factor π^{-1} is removed from the first term on the right hand side.) If α is fixed and $t \to \infty$, then (10) implies (12), and so the assertion (I) follows.

4. Some corollaries of (I), and its proof, will now be deduced. For example, to show that (2) is sufficient for (1) to be of *Grenzpunkt* type, let $\alpha = \frac{1}{2}t$ and $Q(t) = t/\epsilon$, where $\epsilon > 0$ is fixed. Then (14) becomes

$$\int_{\frac{1}{2}t}^{t} ds/\theta'(s) \geqq 2\epsilon\log 2 - (4\epsilon^2 t^{-2})(\theta(t) - \theta(t/2)).$$

Since $\epsilon > 0$ is arbitrary, it is clear from (16) that if

$$(2')\qquad \liminf_{t\to\infty}\,\{N(t) - N(\tfrac{1}{2}t)\}/t^2 < \infty,$$

then

$$\limsup_{t\to\infty}\int_{\frac{1}{2}t}^{t} ds/\theta'(s) > 0,$$

and so (12) holds. Hence (2), or even (2′), is sufficient.

In the same manner, by replacing $Q(t)$ by $1/\epsilon$ and α by $t - 1$ in (14), one proves the following:

COROLLARY 1. *The equation* (1) *is of the Grenzpunkt type if*

$$\liminf_{t\to\infty}\{N(t)-N(t-1)\}<\infty;$$

for instance, if $q(t)\leqq\text{const.}$ *holds on t-intervals which tend to* ∞ *and are of length* 1.

5. It will be seen below (cf. the Lemma, § 7) that the proof of the following assertion implies that (6), (7), (9) are sufficient for (1) to be of the *Grenzpunkt* type.

COROLLARY 2. *Let* $Q(t)>0$ *be defined for large t and satisfy* (7) *and* (9). *Then the assumption*

$$N(t)\leqq\int^t Q(s)\,ds \text{ for large } t, \tag{19}$$

hence, in particular, the assumption

$$\int^t q^+(s)\,ds\leqq t^{-1}\Big(\int^t Q(s)\,ds\Big)^2 \tag{20}$$

is a sufficient condition in order that (1) *be of Grenzpunkt type.*

As an application of Corollary 2, let $Q(t)$ be a constant multiple of $t\log t$. Then

$$N(t)=O(t^2\log t), \tag{2''}$$

or, more particularly,

$$\int^t q^+(s)\,ds=O(t^3\log^2 t), \tag{4''}$$

assures that (1) is of *Grenzpunkt* type.

In the proof of Corollary 2, it is sufficient to consider (19), since (19) is a consequence of (20), by (3) (as the function $Q(s)$ in (19) can be a multiple of that in (20), to account for the $O(1)$ term in (3)). In order that the proof include the case when (6), (7), (9) is assumed (cf. the Lemma, § 7, below), Corollary 2 will be proved with (19) relaxed to

$$N(t)\leqq\int^t Q(s)\,ds+C\log Q(t) \text{ for large } t. \tag{21}$$

(Actually, the $\log Q$ can be improved to $Q^{2-\epsilon}$). In view of Corollary 1, it can be assumed that $Q(s)$ is non-decreasing.

In order to apply (I), let the function $Q(s)$ occurring in (10) be the function $Q(s)$ in (21). If the second integral in the expression $\{\ldots\}$ in (10) is integrated by parts (cf. (17) and (18)), it is seen from $dQ^{-2}(s) \leqq 0$ that $\{\ldots\}$ is not increased if $N(t)$ is replaced by a majorant, say the function on the right-hand side of (21). Hence, $\{\ldots\}$ in (10) is not less than

$$2\int^{t} Q^{-1}(s)\,ds - \int^{t} Q^{-1}(s)\,ds - C\int^{t} Q^{-3}(s)\,dQ(s)$$
$$- Q^{-2}(t) - 2\int^{t} Q^{-3}(s)\,dQ(s),$$

which is $\int^{t} Q^{-1}(s)\,ds + \tfrac{1}{2}CQ^{-2}(t) +$ const. Hence, (10) follows from (7).

6. The fact that (6), (7) and (8) imply that (1) is of *Grenzpunkt* type is contained in the Lemma of § 7, below, and in the following assertion:

COROLLARY 3. *Let* $Q(t) > 0$ *be defined for large* t *and satisfy* (7) *and* (8). *If, in addition,* $Q(t)$ *has the property that*

$$\int_{u}^{v} Q(s)\,ds \geqq 1 \tag{22}$$

whenever $t = u, v$ *are successive zeros of some* ($\not\equiv 0$) *solution of* (1), *then* (1) *is of the Grenzpunkt type.*

It is clear that (19) is a consequence of the condition involving (22), but not conversely. In the proof of Corollary 3, it can be supposed that $Q(t)$ does not tend to 0, as $t \to \infty$; for otherwise $N(t) = o(t)$, and so (1) is *Grenzpunkt* type. Similarly, it can be supposed that $\int^{\infty} Q(s)\,ds = \infty$. Let $t_1 < t_2 < \cdots$ be an unbounded sequence of t-values such that

$$\tfrac{1}{2} \leqq \int_{t_{k-1}}^{t_k} Q(s)\,ds \leqq 1 \tag{23}$$

and

$$\liminf_{k\to\infty} Q^{-2}(t_k) < \infty. \tag{24}$$

By virtue of the second inequality in (23), the hypothesis on $Q(s)$ implies

$$\int_{t_{k-1}}^{t_k} dN(s) \leqq 2. \tag{25}$$

The assumption (8), which is equivalent to the existence of a constant $C > 0$ satisfying

$$Q^{-2}(s)\,|\,dQ(s)\,| \leqq C ds \tag{26}$$

in the sense of Stieltjes' integration, shows that

$$|\log Q(v)/Q(u)\,| \leqq \int_{t_{k-1}}^{t_k} Q^{-1}(s)\,|\,dQ(s)\,| \leqq C \int_{t_{k-1}}^{t_k} Q(s)\,ds \leqq C,$$

if $t_{k-1} \leqq u < v \leqq t_k$. Hence, if $Q(u_k)$ and $Q(v_k)$ are the minimum and maximum values, respectively, of $Q(t)$ for $t_{k-1} \leqq t \leqq t_k$, then there exists a constant K satisfying

$$Q(v_k)/Q(u_k) \leqq K. \tag{27}$$

In order to apply (I), let the function $Q(s)$ in (10) be $1/\epsilon$ times the function $Q(s)$ occurring in (22), where $\epsilon > 0$ is a constant to be specified below, and let $t = t_k$. Then, by virtue (25), the expression $\{\ldots\}$ in (10) is minorized by

$$2\epsilon \int^{t_k} Q^{-1}(s)\,ds - 2\epsilon^2 \sum^{k} Q^{-2}(u_j) - \epsilon^2 Q^{-2}(t_k) - \epsilon^2 \int^{t_k} |\,dQ^{-2}(s)\,|.$$

By the first inequality in (23),

$$Q^{-2}(u_j) \leqq 2Q^{-2}(u_j) \int_{t_{j-1}}^{t_j} Q(s)\,ds; \text{ hence } Q^{-2}(u_j) \leqq 2K^2 \int_{t_{j-1}}^{t_j} Q^{-1}(s)\,ds.$$

by (27). Also, (26) implies that $\int^{t_k} |\,dQ^{-2}(s)\,| \leqq 2C \int^{t_k} Q^{-1}(s)\,ds.$ Thus, the expression $\{\ldots\}$ in (10) is minorized by

$$\epsilon(2 - 4\epsilon K^2 - 2\epsilon C) \int^{t_k} Q^{-1}(s)\,ds - \epsilon^2 Q^{-2}(t_k).$$

Consequently, if $\epsilon > 0$ is sufficiently small, (10) follows from (7) and (24). This proves Corollary 3.

7. In this section, there will be proved a lemma which is of interest in itself and which shows that Corollary 2 (or 3) implies the sufficiency of

(6), (7) and (9) (or (8)). This lemma resulted from a discussion which the author had with Professor Wintner. It furnishes a simple proof of the asymptotic formula for $N(t)$ given in [2].

LEMMA. *Let* $Q(t) > 0$ *be a continuous function of bounded variation on the interval* $0 \leqq t \leqq T$ *and let* n *be the number of zeros of a solution* $z = z(t) \not\equiv 0$ *of*

$$z'' + Q^2 z = 0. \tag{28}$$

Then there exists a number τ *satisfying* $|\tau| \leqq \pi$ *and*

$$\pi n = \int_0^T Q(s)\,ds + \tau\left(\tfrac{1}{2}\int_0^T Q^{-1}(s)\,|dQ(s)| + 2\right). \tag{29}$$

In order to prove the Lemma, let $\theta = \theta(t)$ be the continuous function defined by

$$\theta(t) = \arctan\,(Qz/z'), \qquad |\theta(0)| \leqq \pi. \tag{30}$$

Since $Q > 0$ and since $\theta(t)$ increases at the t-values where $z(t) = 0$, it is clear that

$$|\theta(T) - \pi n| \leqq \pi. \tag{31}$$

From (30),

$$d\theta = (z'^2 + Q^2 z^2)^{-1}\{Q(z'^2 - zz'')\,dt + zz'\,dQ\}.$$

But $z'' = -Q^2 z$, so that

$$d\theta = Q\,dt + (z'^2 + Q^2 z^2)^{-1}(Qzz')\,Q^{-1}dQ.$$

Since $|Qzz'| \leqq \frac{1}{2}(z'^2 + Q^2 z^2)$, the Lemma follows from (31).

8. Although more general criteria for (1) to be of *Grenzpunkt* type result from the methods employed above, than from the method initiated in [9], the latter has the advantage of providing information about the L^2-character of the derivative $x'(t)$ of a solution $x = x(t)$ of (1) which is of class L^2. For example, it was proved in [9] that if $q(t) \leqq$ const. and $x(t)$ is of class L^2 and a solution of (1), then $x'(t)$ is of class L^2. A variant of this theorem will be obtained for the case that $q(t)$ is "nearly" a constant.

(II) *Let* $q(t)$, $Q(t)$ *be real-valued continuous functions for* $0 \leqq t < \infty$. *In addition, let* $Q(t)$ *satisfy* (8),

$$Q(t) \geqq \text{const.} > 0 \tag{32}$$

and

$$\left|\int_0^t q(s)\,ds\right| \leqq Q(t). \tag{33}$$

Let $x = x(t)$ *be of class* L^2 *and a solution of*

$$x'' + (\lambda + q)x = 0, \tag{34}$$

where λ *is a constant. Then* $x'(t)/Q(t)$ *is of class* L^2.

It may be mentioned that, by using a device similar to that in [6], the assumption (8) in (II) can be replaced by (9) if, in the conclusion, $x'(t)/Q(2t)$ replaces $x'(t)/Q(t)$.

The use of (II) shortens the proof of a theorem in [1] which states that if

$$\limsup_{t\to\infty} \int^t |q(s)|\,ds/\log t = \lambda_0^{\frac{1}{2}} < \infty \qquad (\lambda_0 \geqq 0), \tag{35}$$

then (34) cannot possess a ($\not\equiv 0$) solution of class L^2 when $\lambda > \lambda_0$. For suppose that (34) possesses a solution $x = x(t) \not\equiv 0$ of class L^2, then, by (II), the function $x'/\log t$, as well as $x/\log t$, is of class L^2. On the other hand, if $\lambda > 0$, it follows that

$$|(x^2 + x'^2/\lambda)'| = |-2qxx'/\lambda| \leqq (x^2 + x'^2/\lambda)\,|q|/\lambda^{\frac{1}{2}};$$

so that

$$x^2 + x'^2/\lambda \geqq \text{const.} \exp\left(-\int_0^t |q(s)|\,ds/\lambda^{\frac{1}{2}}\right),$$

where const. > 0. This leads to a contradiction in view of (35) and $\lambda > \lambda_0$.

9. In order to prove (II), let (34) be multiplied by $x = x(t)$. The resulting equation can be written in the form $x'^2 = (xx')' + \lambda x^2 + qx^2$. Hence

$$\int_0^t (x'/Q)^2 ds = \int_0^t (xx')'Q^{-2}ds + \lambda \int_0^t x^2Q^{-2}ds + \int_0^t qQ^{-2}x^2ds. \tag{36}$$

After an integration by parts, the first integral on the right-hand side of (36) becomes

$$xx'Q^{-2}\,\big|_0^t + 2\int_0^t xx'Q^{-3}dQ.$$

The assumption (8), that is, (26), means that the last integral is

$$(37)\qquad O\Big(\int_0^t |xx'Q^{-1}|\,ds\Big) = O(1)\Big(\int_0^t x^2ds\Big)^{\frac{1}{2}}\Big(\int_0^t (x'/Q)^2ds\Big)^{\frac{1}{2}},$$

by Schwarz's inequality. Thus the assumption (32) shows that, as $t\rightarrow\infty$,

$$(38)\qquad \int_0^t (xx')'Q^{-1}ds = O(|xx'|) + O\Big(\int_0^t (x'/Q)^2ds\Big)^{\frac{1}{2}} + O(1).$$

The relation (32) also shows that

$$(39)\qquad \lambda\int_0^t x^2Q^{-2}ds = O(1),$$

as $t\rightarrow\infty$. If $I = I(t)$ denotes the integral whose absolute value occurs in (33), then the last integral in (36) is

$$Ix^2Q^{-2}\,|_0^t - 2\int_0^t Ixx'Q^{-2}ds + 2\int_0^t Ix^2Q^{-3}dQ.$$

By (26), (32) and (33), this expression is

$$O(x^2) + O\Big(\int_0^t |xx'Q^{-1}|\,ds\Big) + O\Big(\int_0^t x^2ds\Big);$$

hence, by (37),

$$(40)\qquad \int_0^t qQ^{-2}x^2ds = O(x^2) + O\Big(\int_0^t (x'/Q)\,ds\Big)^{\frac{1}{2}} + O(1).$$

Consequently, (36), (38), (39) and (40) give

$$\int_0^t (x'/Q)^2ds = O(|xx'| + x^2) + O\Big(\int_0^t (x'/Q)^2ds\Big)^{\frac{1}{2}}.$$

Thus the assertion (II) follows if it is ascertained that

$$(41)\qquad \liminf_{t\rightarrow\infty}\,(|xx'| + x^2) < \infty.$$

The relation (41) clear holds if $x(t)$ possesses an infinity of zeros (which then necessarily tend to ∞). Thus it can be supposed that $x(t) > 0$ for large t. It can also be supposed that $x(t)$ is monotone for large t, otherwise the fact that $x(t)$ is of class L^2 implies that $x(t)$ has a sequence of local minima $x(t_n)$ satisfying $x'(t_n) = 0$, $x(t_n) \rightarrow 0$ and $t_n \rightarrow \infty$, as $n \rightarrow \infty$. In which case, (41) holds. But the monotony of $x(t)$ means that

$$\int^t (|xx'| + x^2)ds = |\int^t xx'ds| + \int^t x^2ds = \tfrac{1}{2}x^2 + O(1)$$

has a finite limit inferior, as $t \rightarrow \infty$. Hence (41) holds in any case. This proves (II).

PARIS, FRANCE.

REFERENCES.

[1] G. Borg, "On the point spectrum of $y'' + (\lambda - q(x))y = 0$," *American Journal of Mathemataics*, vol. 73 (1951), pp. 122-126.

[2] P. Hartman and A. Wintner, "The asymptotic arcus variation of solutions of real linear differential equations of second order," *ibid.*, vol. 70 (1948), pp. 1-10.

[3] ——— and A. Wintner, "Criteria of non-degeneracy for the wave equation," *ibid.*, vol. 70 (1948), pp. 295-308.

[4] ——— and A. Wintner, "A criterion for the non-degenracy of the wave equation," *ibid.*, vol. 71 (1949), pp. 206-213.

[5] N. Levison, "Criteria for the limit-point case for second order linear differential operators," *Časopis pro pěstováni matematiky a fysiky*, vol. 74 (1949), pp. 17-20.

[6] D. B. Sears, "Note on the uniqueness of the Green's functions associated with certain differential equations," *Canadian Journal of Mathematics*, vol. 2 (1950), pp. 314-325.

[7] E. C. Titchmarsh, "On the uniqueness of the Green's functions associated with a second order differential equation," *ibid.*, vol. 1 (1949), pp. 191-198.

[8] H. Weyl, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[9] A. Wintner, "(L^2)-connections between the potential and kinetic energies of linear systems," *American Journal of Mathematics*, vol. 69 (1947), pp. 5-13.

[10] ———, "On the normalization of the characteristic differentials in continuous spectra," *Physical Review*, vol. 72 (1947), pp. 516-517.

ON BOUNDED GREEN'S KERNELS FOR SECOND ORDER LINEAR ORDINARY DIFFERENTIAL EQUATIONS.*

By PHILIP HARTMAN.[1]

1. Let $q = q(t)$, where $0 \leqq t < \infty$, be a real-valued continuous function satisfying

$$q(t) < C^2, \tag{1}$$

where $C > 0$ is a constant. Then the differential equation

$$x'' + (\lambda + q)x = 0 \tag{2}$$

and a homogeneous boundary condition at $t = 0$ determine a self-adjoint boundary value problem in the Hilbert space $L^2 = L^2(0, \infty)$; [10], p. 238, Let the set S of λ-values which are cluster points of the spectrum of such a boundary value problem be called the "essential spectrum" of (2). This terminology is justified by the fact that S is independent of the particular boundary condition which defines it; [10], p. 251.

If λ is a (real) value not in S, then (2) possesses a solution $x = z(t) \not\equiv 0$ of a class L^2; [5]. It has been shown by Wintner [13] that such a solution, $z(t)$, and its derivative, $z'(t)$, are majorized by t^{-n} as $t \rightarrow \infty$, for every n, and he has raised the question whether or not the estimate t^{-n} can be replaced by an exponentially small one. This question has been answered in part by Putnam [9], who has shown that such an estimate exists if condition (1) is strengthened to

$$|q(t)| < C^2 \tag{3}$$

and λ is subject to the restriction that $|\lambda + q| \geqq \text{const.} > 0$ for large t. The object of this paper is to answer Wintner's question (without any restrictive hypothesis) in the affirmative.

(*) *If $q = q(t)$ is a continuous function on $0 \leqq t < \infty$ satisfying* (1) *and if λ is a real number not in the essential spectrum of* (2), *then there exists a number $\epsilon = \epsilon(\lambda) > 0$ and a solution $x = z(t)$ of* (2) *satisfying*

$$\limsup_{t \rightarrow \infty} (z^2 + z'^2) e^{\epsilon t} < \infty; \tag{4}$$

---

* Received November 27, 1950.

[1] John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

every solution of (2), $x = y(t)$, *linearly independent of* $x = z(t)$ *satisfies*

$$\liminf_{t\to\infty} (y^2 + y'^2) e^{-\epsilon t} > 0. \tag{5}$$

It can be mentioned that (*) is false if the condition (1) is dropped; [7], p. 648. Also, the following converse of (*) is false ([4,], p. 860): If q satisfies (1), or even (3), and if (2) possesses non-trivial solutions $x = z(t)$, $y(t)$ satisfying (4), (5), respectively, for some $\epsilon > 0$, then λ is not in S.

As a corollary of the proof of (*), see (**) in § **10** below, it will be seen that ϵ is arbitrary in (4) and (5) if, for every λ, the non-trivial solutions of (2) have only a finite number of zeros.

2. In the proof of (*), it can obviously be supposed that

$$C > 1. \tag{6}$$

It can also be supposed that $\lambda = 0$ (for the addition of λ to q translates S by $-\lambda$); in this case, (2) reduces to

$$x'' + qx = 0. \tag{7}$$

It is understood that only real-valued solutions of (7) occur below.

Since $\lambda = 0$ is not in S, the differential equation (7) possesses a non-trivial solution $x = z(t)$ of class L^2; see [5]. Let $x = y(t)$ be any solution of (2) which is linearly independent of $x = z(t)$. Then the integral operator with the Green kernel $G(s,t) = G(t,s) = z(s)y(t)$ for $0 \leqq t \leqq s < \infty$ is bounded in L^2; cf. [6]. That is to say, if $g(t)$ is of class L^2 and

$$h = h_g(t) = z(t)\int_0^t y(s)g(s)\,ds + y(t)\int_t^\infty z(s)g(s)\,ds, \tag{8}$$

then $h(t)$ is of class L^2 and

$$\int_0^\infty h^2 ds \leqq M \int_0^\infty g^2 ds, \tag{9}$$

where M is a constant independent of g.

If $T > 0$ and $g = g_T(t)$ is defined to be 0 or $z(t)$ according as $0 \leqq t \leqq T$ or $T < t < \infty$, then $h_g(t) = y(t)\int_T^\infty z^2 ds$ for $0 \leqq t \leqq T$. Hence (9) implies

$$(10) \qquad \left(\int_0^T y^2 ds\right)\left(\int_T^\infty z^2 ds\right) \leqq M$$

for $0 < T < \infty$.

3. It is clear from the statement of (*) that, without loss of generality, it can be supposed that

$$(11) \qquad y(0) = 0 \text{ or } y'(0) = 0.$$

If (7) is multiplied by x, and xx'' is written in the form $(xx')' - x'^2$, it is seen ([12], p. 9) that (11) and (1) imply

$$\int_0^T y'^2 ds \leqq y(T)y'(T) + C^2 \int_0^T y^2 ds;$$

while (1) implies that $z'(t)$ is of class L^2, $zz' \rightarrow 0$ as $t \rightarrow \infty$, and

$$\int_T^\infty z'^2 ds \leqq -z(T)z'(T) + C^2 \int_T^\infty z^2 ds$$

([12], p. 9). The last two inequalities are strengthened if C is replaced by $2C$, hence

$$(12) \qquad 2\int_0^T y^2 ds \geqq Y(T) - y(T)y'(T)/4C^2$$

and

$$(13) \qquad 2\int_T^\infty z^2 ds \geqq Z(T) + z(T)z'(T)/4C^2,$$

where

$$(14) \qquad Y(T) = \int_0^T (y^2 + (y'/2C)^2 ds \text{ and } Z(T) = \int_T^\infty (z^2 + (z'/2C)^2) ds.$$

Since $x = y(t)$, $z(t)$ are linearly independent solutions of (7), their Wronskian is a non-vanishing constant

$$(15) \qquad yz' - y'z = w \neq 0.$$

Hence, by the Schwarz inequality,

$$(16) \qquad -Z'Y' = (z^2 + (z'/2C)^2)(y^2 + (y'/2C)^2) \geqq w^2/4C^2 > 0.$$

Thus, (10), (12), (13) and (16) imply that the product $P_1(T,U)P_2(T,U)$, where

$$(17)\qquad P_1(T,U)=(Y(T)-y(T)y'(T)/4C^2)/Y'(U)$$

and

$$(18)\qquad P_2(T,U)=(Z(T)+z(T)z'(T)/4C^2)/(-Z'(U)),$$

is bounded from above for $0<T,U<\infty$.

For the purpose of § 9, Appendix, below, it can be mentioned that an upper bound for P_1P_2 is $(M)(4)(4C^2/w^2)$; cf. (10), (12), (13) and (16).

4. The object of the remainder of the proof will be to appraise $P_2(T,U)$ from the above for a suitable choice of $U=U(T)$. To this end, it will be convenient to appraise $P_1(T,U)$ from below. First,

$$y^2(T)-y^2(0)=\int_0^T 2yy'ds\leqq 2C\int_0^T (y^2+(y'/2C)^2)\,ds=2CY(T);$$

so that, for sufficiently large T (since y is not of class L^2),

$$(19)\qquad y^2(T)\leqq 9CY(T)/4\leqq 3CY(T).$$

Also

$$(20)\qquad y'^2(T)-y'^2(0)=\int_0^T 2y'y''ds=-\int_0^T 2qyy'ds.$$

As the case when (3), rather than (1), is assumed is quite simple from this point onward, the proof of (*) will first be indicated for this case. This will make clear the meaning of the computations to follow. By (3) and (20),

$$(21)\quad (y'(T)/2C)^2\leqq \text{Const.}+\tfrac{1}{2}\int_0^T |\,yy'\,|\,ds\leqq(3C/4)\int_0^T (y^2+(y'/2C)^2)ds$$

for sufficiently large T (since y is not of class L^2). Thus, (19) and (21) give $Y'(T)\leqq 3CY(T)$ for large T. Also,

$$(22)\qquad |\,yy'/4C^2\,|\leqq(y^2+(y'/2C)^2)/4C.$$

Thus, by (19), (21) and (22), it follows that $P_1(T,T)$ is bounded from below, for large T, by $1/3C-1/4C>0$. (This inequality was the reason for replacing C by $2C$ in the inequality preceding (12).) Consequently, $P_2(T, T)$ is bounded from above. The inequality (22), where y is replaced by z, shows that $Z(T)/(-Z'(T))$ is also bounded from above, say by $1/\epsilon$. The proof can now be completed as in [9], pp. 146-147. The differential inequality for Z implies

$$(23) \qquad Z(T) \leqq Ke^{-\epsilon T} \qquad (K = \text{const.})$$

for $0 \leqq T < \infty$. Since $z(\infty) = 0$ ([12], p. 8),

$$(24) \qquad z^2(T) = -\int_T^\infty 2zz'ds \leqq 2CZ(T),$$

(cf. formula proceeding (19)) and since $z'(\infty) = 0$ ([12], p. 13),

$$(25) \qquad z'^2(T) \leqq 4C^3Z(T)$$

(cf. (20) and (21)). Hence (4) follows from (23), (24) and (25). Then (5) follows from (15) and (16). This proves (*) for the case that (3) is assumed.

For the purpose of § **9**, below, it can be mentioned that the number $1/\epsilon$ above is $(M)(4)(4C^2/w^2)(1/3C - 1/4C)^{-1} + 1/4C$; cf. the end of § **3** and the three sentences following (22).

5. In the general case, where (1) is assumed, the estimates (21), (25) are not available for the derivatives $y'(T)$, $z'(T)$. A substitute is furnished by a Tauberian type of argument adapted from [1].

LEMMA. *Let* $q = q(t)$ *satisfy the conditions of* (*), *let* $x = x(t) \not\equiv 0$ *be a solution of* (7) *and* $2/C \leqq T < \infty$. *If* $x'(T) \neq 0$, *let* $T^* = T^*(T)$ *be defined as the largest (smallest) possible value satisfying both* $|T^* - T| \leqq 2/C$ *and* $xx' \geqq 0$ $(\leqq 0)$ *for* $T \leqq t \leqq T^*$ $(T^* \leqq t \leqq T)$. *Then*

$$(26) \qquad x'^2(T) \leqq 4C^2x^2(T^*).$$

Proof of the Lemma. Consider the case $xx' \geqq 0$ for $T \leqq t \leqq T^*$. Since x and x' cannot vanish simultaneously, xx' changes sign at a zero of x. Hence x, and therefore, x' do not change sign on the interval (T, T^*). It can be supposed that $x \geqq 0$ and $x' \geqq 0$ (otherwise x is replaced by $-x$). Thus $0 < x(t) \leqq x(T^*)$ for $T < t \leqq T^*$. By (7) and (1), it is seen that $x''(t) \geqq -C^2x(T^*)$ or $x'(t) \geqq -C^2x(T^*)(t - T) + x'(T)$. If (26) does not hold, then $x'(t) > -C^2x(T^*)(t-T) + 2Cx(T^*)$. The right-hand side of the last inequality is non-negative for $T \leqq t \leqq T + 2/C$; so that, by definition, $T^* = T + 2/C$. Another integration shows that $x(T^*) > -2x(T^*) + 4x(T^*)$, since $x(T) \geqq 0$. But this leads to the contradiction $x(T^*) > 2x(T^*) > 0$, and proves the Lemma for this case. The proof for the case $xx' \leqq 0$ is similar.

6. It will be convenient to simplify the factor $P_1(T, U)$ of the product P_1P_2. It will be shown that

(27) $\quad (Y(T)/Y'(U))P_2(T, U)$ is bounded from above

for $0 < T, U < \infty$. To this end, the linear transformation $g \to h$ in (8) will be reconsidered. For an arbitrary g of class L^2, the function h possesses a derivative

$$h' = z'(t) \int_0^t y(s)g(s)\,ds + y'(t) \int_t^\infty z(s)g(s)\,ds, \tag{28}$$

which is absolutely continuous. The second derivative of h satisfies (almost everywhere) the inhomogeneous differential equation

$$h'' + qh = wg,$$

where w is the Wronskian constant (15). Furthermore, $h(0) = 0$ or $h'(0) = 0$ by (8) or (28) and (11); also, $hh' \to 0$ as $t \to \infty$; and so

$$\int_0^\infty h'^2\,ds \leqq C^2 \int_0^\infty h^2\,ds + |w| \int_0^\infty |gh|\,ds$$

([12], p. 9; cf. the derivation of the inequality preceding (12)). Consequently, (9) implies that the linear transformation $g \to h'$ is bounded,

$$\int_0^\infty h'^2\,ds \leqq (C^2M + |w|\,M^{\frac{1}{2}}) \int_0^\infty g^2\,ds. \tag{29}$$

If the function $g = g_T(t)$ is chosen as in the derivation of (10), it is seen that

$$\left(\int_0^T y'^2\,ds\right)\left(\int_T^\infty z^2\,ds\right) \leqq C^2M + |w|\,M^{\frac{1}{2}}. \tag{30}$$

Hence, (27) follows from (10), (30), (13) and (16).

For the purposes of § **10**, Appendix, below, it can be mentioned that an upper bound for the function in (27) is $(2M + | w | M^{\frac{1}{2}}C^{-2}) . 2 . (4C^2/w^2)$.

7. The Lemma and the inequality (19) show that if $T \geqq 2/C$, then either $y'(T) = 0$ or $(y'(T)/2C)^2 \leqq 3CY(T^*)$. Since $Y(T)$ is increasing and $T^* \leqq T + 2/C$, it follows that $(y'(T)/2C)^2 < 3CY(T + 2/C)$ in either case. Also, by (19) for sufficiently large T, $y^2(T) \leqq 3CY(T) < 3CY(T + 2/C)$. Consequently, $Y(T + 2/C)/Y'(T)$ is bounded below by $1/6C$ for large T-values. Hence, by (27), the function $P_2(T + 2/C, T)$ is bounded from above, say by α.

For the purposes of § **10**, below, it can be mentioned that α can be chosen to be $(2M + | w | M^{\frac{1}{2}}C^{-2}) (2) (4C^2/w^2) (6C)$; cf. the end of § **6**.

The inequality (22), in which y is replaced by z, gives $| zz'/4C^2 | \leqq - Z'/4C$. Thus, by the definition of $P_2(T, U)$, it follows that Z satisfies the functional-differential inequality

$$(31) \qquad Z(T + 2/C) + Z'(T + 2/C)/4C \leqq - \alpha Z'(T)$$

for $0 < T < \infty$. It will be shown that (31), together with the inequalities $Z \geqq 0$, $Z' \leqq 0$ imply the existence of constants $\epsilon > 0$, K satisfying (23).

Let $0 \leqq U < T$ and let (31) be integrated, with respect to T, over the interval $[U, T]$. One obtains

$$\alpha Z(T) + Z(T + 2/C)/4C \leqq \alpha Z(U) + Z(U + 2/C)/4C - \int_U^T Z(s + 2/C) ds.$$

By the monotony of $Z(T)$,

$$(\alpha + 1/4C) Z(T + 2/C) \leqq (\alpha + 1/4C) Z(U) - \int_U^T Z(s + 2/C) ds.$$

If the abbreviations

$$(32) \qquad \alpha + 1/4C = 1/\beta \text{ and } 2/C = \delta$$

are introduced, the last inequality can be written in the form

$$Z(T + \delta) \leqq Z(U) - \beta \int_U^T Z(s + \delta) ds.$$

If $[U, T]$ becomes $[T, T + \delta]$, then $Z(T + 2\delta) \leqq Z(T) - \beta\delta Z(T + 2\delta)$, by virtue of the monotony of Z. The last inequality is equivalent to $Z(T + 2\delta) \leqq Z(T)(1 + \beta\delta)^{-1}$. For a non-negative integer n, let $T = 2n\delta$.

Then $Z(2n\delta+2\delta)\leqq Z(2n\delta)(1+\beta\delta)^{-1}$; so that, by induction, $Z(2n\delta)\leqq Z(0)(1+\beta\delta)^{-n}$ for $n=0,1,\cdots$. Hence, the monotony of Z implies (23) for $0\leqq T<\infty$, where $K=Z(0)(1+\beta\delta)$ and

$$\epsilon=(1/2\delta)\log(1+\beta\delta)>0. \tag{33}$$

Consequently, (23) and (24) show that

$$z^2(T)\leqq 2CKe^{-\epsilon T}$$

for $0\leqq T<\infty$. By the Lemma, if $T=2/C$, then either $z'(T)=0$ or (26) holds. But since $T^*\geqq T-2/C$, one has

$$z'^2(T)\leqq(8C^3Ke^{2\epsilon/C})e^{-\epsilon T}$$

in either case. Hence (4) holds, while (5) is a consequence of (16) and (6). This completes the proof of (*).

8. *Remark.* When (3) is assumed, one can show that, not only does (4) and (5) hold but, also there exists an $\eta>0$ satisfying

$$\liminf_{t\to\infty}(z^2+z'^2)e^{\eta t}>0 \text{ and } \limsup_{t\to\infty}(y^2+y'^2)e^{-\eta t}<\infty. \tag{34}$$

Thus, for large t,

$$e^{-\eta t}<z^2+z'^2<e^{-\epsilon t} \text{ and } e^{\epsilon t}<y^2+y'^2<e^{\eta t}$$

for suitably chosen positive ϵ and η.

The relations (34) are known ([11], p. 391) and do not depend on the assumption that λ is not in S. They also are a consequence of the fact that $Y(T)/Y'(T)$ is bounded from below for large T (cf. (19) and (21)).

The relations (34) need not hold if (3) is relaxed to (1); cf. § **10**, below. This is shown by the example in which $q(t)=2-4t^2$. Since $q(t)\to-\infty$, as $t\to\infty$, the set S is empty; [10], p. 238. On the other hand, the case $\lambda=0$ of (2), that is, (7) has the solution $x=z(t)=\exp(-t^2)$. Hence, the first relation in (34) is violated for every η. In particular, $Y(T)/Y'(T)$ cannot be bounded from below for large T, although if C is chosen to be 2, then $Y(T+1)/Y'(T)$ is (cf. § **7**, above).

Appendix.

9. The essential spectrum S is a closed set; so that, its complement is an open set. The intervals on the real λ-axis occurring in the canonical decomposition of this open set will be called the gaps in the essential spectrum of (2). For a given λ not in S, let $L(\lambda)$ denote the least upper bound of the set of ϵ satisfying (4) and (5). (Since $z = z(t)$ is unique up to a constant factor, it is clear that L is determined by λ and not by any choice of solutions of (2).) In a recent paper [4], one of the problems considered was an upper estimate for $L(\lambda)$, as $\lambda \rightarrow \infty$ (on the gaps of the essential spectrum, if any) in the case that q satisfies (3). Since the above proof for the existence of an $\epsilon = \epsilon(\lambda)$ is "constructive," it provides a lower estimate for $L(\lambda)$. It can be expected that the problem of obtaining lower estimates for $L(\lambda)$ can be quite delicate for, as suggested by Wintner, if λ is on a gap and tends to an endpoint, then $L(\lambda)$ tends to 0, while $L(\lambda)$ attains its "maximum" on the gap when λ is near the midpoint. Such a behavior is indicated by the lower estimate for $L(\lambda)$ provided by the above proof.

For the case when $\lambda = 0$ and (3) holds, it is seen, by the end of § **4**, that (4) and (5) hold if $\epsilon = (AC^3M + B/C)^{-1}$, where A and B are absolute constants and M is the square of the norm of the Green integral operator (8) (when the Wronskian constant w in (15) is 1).

When λ is arbitrary, C^2 can be replaced by $\lambda + C^2$; so that for large λ, the "C" in the formula for ϵ is asymptotically equal to $\lambda^{\frac{1}{2}}$. Also, $M = M(\lambda)$ is the reciprocal of the square of the distance from λ to the nearest point of the spectrum belonging to the homogeneous boundary condition satisfied by $x = y(t)$ at $t = 0$. If λ is fixed, it is seen from the characterization of the spectrum given in [3] that if the interval $[0, \infty)$ is replaced by $[T, \infty)$ and $T \rightarrow \infty$, then $M(\lambda) = M_T(\lambda) \rightarrow D^{-2}(\lambda)$, where $D(\lambda)$ is the distance from λ to the nearest point of the essential spectrum. (This limit process clearly does not affect $L(\lambda)$.) Finally, (3) implies that $D(\lambda)$ is bounded, as $\lambda \rightarrow \infty$ ([8]; cf. also [4]); so that the term $B/C \sim B/\lambda^{\frac{1}{2}}$ is small when compared to $AC^3M \sim A\lambda^{3/2}D^{-2}(\lambda)$, as $\lambda \rightarrow \infty$.

Summarizing, if (3) holds, one obtains the lower estimate Const. $D^2(\lambda)/\lambda^{3/2}$ for $L(\lambda)$ when λ is large (and on a gap of the essential spectrum, if any). This contrasts with the upper estimate Const. $/\lambda^{\frac{1}{2}}$ obtained in [4].

10. In the case that (1) is assumed, it is seen that the ϵ constructed in § **7** is of the form $\epsilon = (C/4) \log \{1 + 8(AC^4M + BC^2M^{\frac{1}{2}} + 1)^{-1}\}$, for the case when $\lambda = 0$ and C, M have the same significance as in the preceding section, while A, B are absolute constants; cf. (32), (33) and the value for α given in the paragraph preceding (31). It can be remarked that if C^2M is large, as in the preceding section, then ϵ is of the same order of magnitude, $1/C^3M$, as obtained above. But our interest in this section lies in the opposite direction.

As above, if the half-axis $[0,\infty)$ is replaced by $[T,\infty)$, it is seen that, when the essential spectrum S is not empty, then $M = M_T \rightarrow D^{-2}(0)$, as $T \rightarrow \infty$, where $D(0)$ is the distance from $\lambda = 0$ to the nearest point of S. But if S is empty (and so, $D(0) = \infty$), it is seen that $M_T \rightarrow 0$, as $T \rightarrow \infty$. Thus, in this case, one can choose, for example, $\epsilon > \frac{1}{2}(C/4)\log(1+8)$. But the constant C can be chosen arbitrarily large (since it is only subject to the restriction (1)). Hence, in (4) and (5), the number ϵ can be chosen arbitrarily large. Since the essential spectrum is empty, whenever (11) is non-oscillatory for every λ ([2], p. 707), the following theorem is proved:

(**) *Let $q = q(t)$, where $0 \leqq t < \infty$, be continuous and satisfy* (1) *for some constant $C > 0$. In addition, let q have the property that for every λ, where $-\infty < \lambda < \infty$, the non-trivial solutions of* (2) *have only a finite number of zeros for $0 \leqq t < \infty$. Then* (2) *has a solution $x = z(t) \not\equiv 0$ satisfying* (4) *for every ϵ; any solution of* (2), *$x = y(t)$, linearly independent of $x = z(t)$, satifies* (5) *for every ϵ.*

This theorem is trivial if (2) is non-oscillatory for every λ by virtue of the condition $q \rightarrow -\infty$, as $t \rightarrow \infty$; see, for example, the comparison theorem in [7], p. 635. But (2) can be non-oscillatory for every λ when the limit superior of q, as $t \rightarrow \infty$, is positive or even $+\infty$; cf. [2], p. 708. On the other hand, if (2) is non-oscillatory for every λ, but q fails to satisfy (1), then (2) need not have a solution which tends to 0 or is even bounded, as $t \rightarrow \infty$; [7], p. 648.

PARIS, FRANCE.

REFERENCES.

[1] P. Hartman, " The L^2-solutions of linear differential equations of second order," *Duke Mathematical Journal*, vol. 14 (1947), pp. 323-326.

[2] ———, " Differential equations with non-oscillatory eigenfunctions," *ibid.*, vol. 15 (1948), pp. 697-709.

[3] ———, "A characterization of the spectra of one-dimensional wave equations," *American Journal of Mathematics*, vol. 71 (1949), pp. 915-920.

[4] ——— and C. R. Putnam, " The gaps in the essential spectra of wave equations," *American Journal of Mathematics*, vol. 72 (1950), pp. 849-862.

[5] ——— and A. Wintner, "An oscillation theorem for continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 33 (1947), pp. 376-379.

[6] ——— and A. Wintner, " On the orientation of unilateral spectra," *American Journal of Mathematics*, vol. 70 (1948), pp. 309-316.

[7] ——— and A. Wintner, " Oscillatory and non-oscillatory linear differential equations," *ibid.*, vol. 71 (1949), pp. 627-649.

[8] C. R. Putnam, " The cluster spectra of bounded potentials," *ibid.*, vol. 70 (1948), pp. 842-848.

[9] ———, " On isolated eigenfunctions associated with bounded potentials," *ibid.*, vol. 72 (1950), pp. 135-147.

[10] H. Weyl, " Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürliche Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 222-269.

[11] A. Wintner, " The adiabatic linear oscillator," *American Journal of Mathematics*, vol. 68 (1946), pp. 385-397.

[12] ———, "(L^2)-connections between the potential and kinetic energies of linear systems," *ibid.*, vol. 69 (1947), pp. 5-13.

[13] ———, " On the smallness of isolated eigenfunctions," *ibid.*, vol. 71 (1949), pp. 603-611.

ON THE EIGENVALUES OF DIFFERENTIAL EQUATIONS.*

By PHILIP HARTMAN.[1]

In the differential equation

$$x'' + (\lambda - q)x = 0, \tag{1}$$

let λ be a real parameter and $q = q(t)$, $0 \leqq t < \infty$, a real-valued continuous function satisfying

$$q(t) \to \infty, \text{ as } t \to \infty. \tag{2}$$

Then (1) and a real homogeneous boundary condition

$$x(0) \sin \theta - x'(0) \cos \theta = 0 \tag{3}$$

determine on $L^2(0, \infty)$ a self-adjoint boundary value problem, the spectrum of which is an unbounded sequence of numbers $\lambda_0 < \lambda_1 < \cdot \cdot \cdot$, and the eigenfunctions belonging to $\lambda = \lambda_n$ have exactly n zeros on $0 < t < \infty$ (Weyl [6]).

It has been proved by Milne [4] that if q possesses a continuous third derivative and if $q' > 0$, $q'' \geqq 0$ and $q'' = o(q'^{4/3})$, as $t \to \infty$, then, as $n \to \infty$,

$$\pi n \sim \int_0^r (\lambda_n - q)^{\frac{1}{2}} dt, \text{ where } r = r_n \tag{4}$$

is the t-value satisfying $q(t) = \lambda_n$. The assumption of the existence of a third derivative, as pointed out by Milne, was unessential to his proof. Titchmarsh [5] gave a simplified proof of (4) retaining all of Milne's conditions, except that concerning the existence of a third derivative. It has been shown by Wintner and the author [3] that (4) holds under the mere assumption that the graph of $q = q(t)$ is convex upwards; for example, if $q(t)$ has a non-negative second derivative. It will be seen below, however, that the assumption of convexity in all of these proof is quite artificial. In fact, convexity implies that the first derivative $q'(t)$ exists, except possibly

---

* Received October 31, 1950.

[1] John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

for a denumerable set of t-values, and that $q'(t) \geqq \text{const.} > 0$ for large t. This inequality, or even

$$(5) \qquad q'(t)t^3 \to \infty, \text{ as } t \to \infty,$$

will be seen to be sufficient for (4). On the other hand, in the sufficient condition (5), the exponent 3 cannot be replaced by a larger exponent, $3 + \epsilon$; in fact, if $q(t)$ satisfies (2) and possesses a continuous positive derivative for which

$$(6) \qquad 0 < \liminf_{t\to\infty} q'(t)t^3 < \infty$$

holds, then (4) need not hold.

The proof of (4) in [3] depends on the method used in [2] to derive an asymptotic formula for the number of zeros on $(0, t)$ of a solution of a differential equation of the type

$$(7) \qquad x'' + Q(t)x = 0,$$

where no parameter λ occurs. It turns out that the asymptotic formula in question is a simple consequence of the following lemma of Wintner and the author, see [1]:

LEMMA. *Let $Q(t) > 0$ be a continuous function of bounded variation on $0 \leqq t \leqq T$ and let M be the number of zeros of a solution $x = x(t) \not\equiv 0$ of (7) for $0 < t < T$. Then*

$$(8) \qquad \left| \pi M - \int_0^T Q^{\frac{1}{2}}(s)\,ds \right| \leqq \tfrac{1}{4} \int_0^T Q^{-1}(s) \,|\, dQ(s) \,| + 2\pi.$$

Correspondingly, with the aid of this lemma, it will be proved that

(i) *the asymptotic formula* (4) *holds if $q(t)$ is an increasing function satisfying, as $t \to \infty$,*

$$(9) \qquad X(t) = \operatorname*{g.\,l.\,b.}_{t \leqq u < v < \infty} (q(v) - q(w)) \Big/ \int_u^v s^{-3}\,ds \to \infty$$

(for example, if $q(t)$ has a continuous derivative satisfying (5)*),*

(ii) *but* (4) *need not hold if $q(t)$ has a continuous derivative satisfying* (6).

Condition (5), or correspondingly (9), is not a condition on the rate of growth of $q(t)$ but a condition on the regularity of growth of the function

$t = r(\lambda)$ which is inverse to $\lambda = q(t)$. Thus, in terms of $r(\lambda)$, condition (5) is

(5 bis) $$r' = o(r^3), \text{ as } \lambda \to \infty \qquad (r' = dr/d\lambda),$$

while (9) is

(9 bis) $$Y(\lambda) = \operatorname*{l.\,u.\,b.}_{\lambda \leqq \mu < \nu < \infty} (r(\nu) - r(\mu)) \Big/ \int_{\mu}^{\nu} r^3(\sigma)\,d\sigma \to 0,$$

as $\lambda \to \infty$.

Proof of (i). After a suitable translation of the λ-axis, it can be supposed that $q(0) = 0$ (and so $r(0) = 0$). Let $N = N(\lambda)$ be the number of zeros on $0 < t < r(\lambda)$ of a solution $x = x(t) \not\equiv 0$ of (1) and (3). Since the coefficient function, $\lambda - q$, of (6) is negative for $t > r(\lambda)$, it follows that $N(\lambda_n)$ is either n or $n - 1$. Hence, (4) follows if it is proved that, as $\lambda \to \infty$,

(10) $$\pi N(\lambda) \sim \int_0^{r(\lambda)} (\lambda - q)^{\frac{1}{2}} dt.$$

For a $\mu = \mu(\lambda)$ satisfying $0 < \mu < \lambda$, let $M = M(\mu, \lambda)$ be the number of zeros of $x = x(t)$ on $0 < t < r(\mu)$. If the Lemma is applied to $Q(t) = \lambda - q(t)$ and $T = r(\mu)$, then, since $|\,dQ\,| = dq$, it follows that

(11) $$\pi M = \int_0^{r(\mu)} (\lambda - q)^{\frac{1}{2}} dt + O(\log \lambda - \log(\lambda - \mu)) + O(1),$$

where the constant implicit in the O-terms are independent of λ and μ. For later purposes, it should be remarked that the term $O(\log \lambda)$ in (11) is unimportant, since the expression on the right-hand side of (10) obviously exceeds const. $r(\frac{1}{2}\lambda)\lambda^{\frac{1}{2}} >$ const. $\lambda^{\frac{1}{2}}$.

On the interval $r(\mu) \leqq t \leqq r(\lambda)$, the coefficient function $\lambda - q$ of (1) is not greater than $\lambda - \mu$. Hence, by the Sturm comparison theorem, the number, $N - M$, of the zeros of the solution $x = x(t)$ on this interval satisfies

(12) $$\pi(N - M) \leqq (r(\lambda) - r(\mu))(\lambda - \mu)^{\frac{1}{2}} + \pi.$$

It is also clear that

(13) $$\int_{r(\mu)}^{r(\lambda)} (\lambda - q)^{\frac{1}{2}} dt \leqq (r(\lambda) - r(\mu))(\lambda - \mu)^{\frac{1}{2}}.$$

On the other hand,

$$(14)\qquad \int_0^{r(\lambda)} (\lambda - q)^{\frac{1}{2}} dt \geqq \int_0^{r(\mu)} (\lambda - q)^{\frac{1}{2}} dt \geqq r(\mu)\,(\lambda - \mu)^{\frac{1}{2}}.$$

Hence, if $\mu = \mu(\lambda)$ satisfies, as $\lambda \rightarrow \infty$,

$$(15)\qquad r(\lambda) \sim r(\mu),$$

the expressions on the left-hand sides of (12) and (13) are equal to an $o(1)$-term times the integral in (10). Thus, by (11) and the remark following it,

$$(16)\qquad N(\lambda) = (1 + o(1)) \int_0^{r(\lambda)} (\lambda - q)^{\frac{1}{2}} dt + O(|\log(\lambda - \mu)|).$$

Consequently, assertion (i) will be proved if it is shown that $\mu = \mu(\lambda)$ can be chosen so as to satisfy (15) and

$$(17)\qquad \log(\lambda - \mu) = o\left(\int_0^{r(\lambda)} (\lambda - q)^{\frac{1}{2}} dt\right), \text{ as } \lambda \rightarrow \infty.$$

To this end, let

$$(18)\qquad \mu = \mu(\lambda) = \lambda - r^{-2}(\lambda);$$

so that, in particular,

$$(19)\qquad \mu \rightarrow \infty, \text{ as } \lambda \rightarrow \infty.$$

Then, by the definition of Y in (9 bis),

$$r(\lambda) \leqq r(\mu) + Y(\mu) \int_\mu^\lambda r^3(\sigma)\, d\sigma.$$

Hence, by (18) and the monotony of $r(\lambda)$, it follows that $r(\lambda) \leqq r(\mu) + Y(\mu) r(\lambda)$. Thus $1 \leqq r(\lambda)/r(\mu) \leqq (1 - Y(\mu))^{-1}$. Consequently, (15) follows from (9 bis) and (19).

The definition of X in (9) implies that

$$\lambda - q(t) \geqq X(t) \int_t^{r(\lambda)} s^{-3} ds = \tfrac{1}{2} X(t)\,(t^{-2} - r^{-2}(\lambda)).$$

If $0 < t \leqq \frac{1}{2} r(\lambda)$, then $(t^{-2} - r^{-2}(\lambda)) \geqq 3t^{-2}/4$. Consequently, for

$$r^{\frac{1}{2}}(\lambda) \leqq t \leqq \tfrac{1}{2} r(\lambda), \text{ one has } \lambda - q(t) \geqq (3/8) X(r^{\frac{1}{2}}(\lambda))\, t^{-2};$$

so that the integral in (17) exceeds $\frac{1}{2}X^{\frac{1}{2}}(r^{\frac{1}{2}}(\lambda))\log\frac{1}{2}r^{\frac{1}{2}}(\lambda)$. On the other hand, the term on the left-hand side of (17) is $\log r^{2}(\lambda)$, by (18). Hence, (17) follows from (9). This completes the proof of (i).

Proof of (ii). This proof is similar to the construction of the counter-example in [2]. By an induction (to be indicated below), define an increasing sequence of positive numbers $\alpha_1, \alpha_2, \cdots$ which tend rapidly to ∞. It is then possible to define an increasing, unbounded function $q = q(t)$ which, on the interval $1 + \alpha_k \leqq t \leqq \alpha_{k+1}$, is equal to $k - t^{-2}$ and which has a continuous derivative satisfying $q'(t)t^3 \geqq 2$ for all large t. It will be seen that if α_{k+1} is chosen so that $\alpha_{k+1} - \alpha_k$ is sufficiently large with respect to α_k, then (4) cannot hold.

If a and η are constants, the solutions of $x'' + t^{-2}x = 0$ are of the form $x = x(t) = at^{\frac{1}{2}}\cos((3^{\frac{1}{2}}/2)\log t + \eta)$ and have therefore an infinity of zeros for $1 \leqq t < \infty$. Hence, if $\alpha_{k+1} - \alpha_k$ is sufficiently large, a solution of (1), when $\lambda = k$, has an arbitrarily large number of zeros on the interval $0 < t < \alpha_{k+1}$. Thus, since an eigenfunction belonging to $\lambda = \lambda_n$ has exactly n zeros, it follows that if $\alpha_{k+1} - \alpha_k$ is sufficiently large, the boundary value problem (1), (3) has an eigenvalue $\lambda = \lambda_n$ arbitrarily near, but less than, k. Let $\beta_k = (1 + \alpha_k)\exp(k\alpha_k)$ and let α_{k+1} be so large that there exists an eigenvalue $\lambda = \lambda_n$ satisfying $k - \beta_k^{-2} \leqq \lambda_n < k - \alpha_{k+1}^{-2}$. In particular, $\alpha_{k+1} > r(\lambda_n) \geqq \beta_k$.

In order to appraise the number $N(\lambda_n)$, consider the Sturm majorant of $\lambda_n - q(t)$ which is equal to k for $0 \leqq t \leqq 1 + \alpha_k$ and to t^{-2} for $1 + \alpha_k \leqq t \leqq r(\lambda_n)$. It follows that

$$\pi N(\lambda_n) \leqq k^{\frac{1}{2}}(1 + \alpha_k) + (3^{\frac{1}{2}}/2)\log\{r(\lambda_n)/(1 + \alpha_k)\} + 4\pi.$$

Since $r(\lambda_n) \geqq \beta_n$, it is seen that $\log\{r(\lambda_n)/(1 + \alpha_k)\} \geqq k\alpha_k$. Thus the first term on the right-hand side in the last formula line is negligible, that is,

$$\pi N(\lambda_n) \leqq (1 + o(1))(3^{\frac{1}{2}}/2)\log\{r(\lambda_n)/(1 + \alpha_k)\}.$$

On the other hand,

$$\int_0^{r(\lambda_n)} (\lambda_n - q)^{\frac{1}{2}}dt \geqq \int_{1+\alpha_k}^{r(\lambda_n)} t^{-1}dt = \log\{r(\lambda_n)/(1 + \alpha_k)\}.$$

Consequently,

$$\liminf_{n\to\infty} \pi N(\lambda_n) \Big/ \int_0^{r(\lambda_n)} (\lambda_n - q)^{\frac{1}{2}}dt \leqq 3^{\frac{1}{2}}/2 < 1.$$

This completes the proof of (ii).

PARIS, FRANCE.

REFERENCES.

[1] P. Hartman, "The number of L^2-solutions of $x'' + q(t)x = 0$," *American Journal of Mathematics*, vol. 73 (1951), pp. 635-645.

[2] ——— and A. Wintner, "The asymptotic arcus variation of solutions of linear differential equations of second order," *ibid.*, vol. 70 (1948), pp. 1-10.

[3] ——— and A. Wintner, "On the asymptotic problems of the zeros in wave mechanics," *ibid.*, vol. 70 (1948), pp. 461-480; see Appendix, pp. 475-480.

[4] W. E. Milne, "On the degree of convergence of expansions in an infinite interval," *Transactions of the American Mathematical Society*, vol. 31 (1929), pp. 906-918.

[5] E. C. Titchmarsh, *Eigenfunction Expansions*, Oxford, 1946; see Chap. VII.

[6] H. Weyl, "Ueber gewöhnliche lineare Differentialgleichungen mit singulären Stellen und ihre Eigenfunctionen," *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, 1909, pp. 37-63.

PARTITIONS AND THEIR REPRESENTATIVE GRAPHS.*

By James K. Senior.

Part I. Introduction.

The phenomenon of isomerism among organic chemical compounds raises some questions which may be answered in part by the solution of certain problems falling within the scope of the mathematical theory of graphs. In the present paper, however, the chemical aspects of these problems are mentioned only in passing; it is intended to treat them more fully in another publication. Most of the terminology hereafter employed is derived from the book by Dénes König entitled *Theorie der endlichen und unendlichen Graphen* (Leipzig, 1936). But, for the sake of brevity, the unmodified word "graph" is here applied only to those diagrams which are more fully described as *finite undirected* graphs. Such a graph may or may not contain loops, and it may or may not be connected.

Consider a finite set P of finite positive integers $p_1, p_2, \cdots p_n$ which (solely for the sake of convenience) are arranged in non-increasing order so that $p_1 \geqq p_2 \geqq \cdots \geqq p_n$. Every such set of integers may be regarded as a partition of the integer ΣP where $\Sigma P = p_1 + p_2 + \cdots + p_n$. It will therefore hereafter be called a partition, with the understanding that ΣP is the integer of which P is the partition. Note that, although $P_a = P_b$ always implies $\Sigma P_a = \Sigma P_b$, the converse statement is true only in the trivial case where $\Sigma P = 1$.

Consider also a finite set C of points $c_1, c_2, \cdots, c_n$ and a finite set of segments called *edges* drawn between these points. If and only if an edge is terminated at both ends by the same point, that point will be called a *loop-point* and that edge will be called a *loop-edge.* The two together constitute a *loop.* A loop-edge will be said to *depend* from its uniquely-determined loop-point. A loop-edge can be part of only one loop, but a loop-point may have an indefinite number of dependent loop-edges, and may besides be one terminal point of an indefinite number of non-loop edges.

The *valence* (ramification order) of a point which terminates r non-loop edges and s loop-edges is defined as $r + 2s$. For any point, either r or s but not both may be zero. If edges are drawn between the points of C in such a

---

* Received August 9, 1949; revised September 10, 1950.

way that each point c_i has the valence p_i, then the graph G thus formed is called a *representative of* P or a *graph for* P; it may be indicated by the symbol $G(P)$. The graph G' is considered to be identical with the above graph G if and only if $n = n'$, and the points (vertices) of G' can be so labelled that:

(1) For every pair of points (c_i', c_j') in G', the number of edges joining c_i' directly to c_j' is the same as the number of edges in G joining c_i directly to c_j.

(2) For every point c_i' in G', the number of loop-edges dependent from c_i' is equal to the number of loop-edges dependent from c_i in G.[1]

Every graph uniquely determines the partition which it represents, for that partition is merely an ordered list of the valences of the points (vertices) of the graph; but not every partition is represented by just one graph. That is to say, the graphs G and G' cannot be identical unless they are graphs for the same P; but they may well be graphs for the same P without being identical.

Let $Z(P)$ be the number of distinct graphs for P.
$ZC(P)$ be the number of distinct connected graphs for P.
$ZL(P)$ be the number of distinct loopless graphs for P.
$ZCL(P)$ be the number of distinct connected loopless graphs for P.

Each of these four non-negative integers is uniquely determined by P.

Evidently (a) If $Z(P) = 0$; then $ZC(P) = ZL(P) = ZCL(P) = 0$
(b) If $ZC(P) = 0$ or $ZL(P) = 0$, then $ZCL(P) = 0$.

It It will hereafter be shown that

(c) If and only if $Z(P) > 0$, then $ZC(P) > 0$ or $ZL(P) > 0$
(d) If and only if $ZC(P) > 0$ and $ZL(P) > 0$, then $ZCL(P) > 0$.

But, for the present paper, the question of chief interest is: When is $ZCL(P)$ less than, when it is equal to, and when is it greater than one?

Although every P uniquely determines its $ZCL(P)$, a given non-negative integer taken as $ZCL(P)$ does not usually (in fact, not in any known case) uniquely determine a corresponding P. Hence there arise two converse problems:

---

[1] There must be x ($\geqq 0$) different ways to label the points (vertices) of G' so as to establish the identity of G and G'. The only question here at issue is: Is $x = 0$, in which case G and G' are distinct; or is $x > 0$, in which case G and G' are identical? Otherwise, the value of x is immaterial for the present discussion.

(A) For any given P, determine the unique value of $ZCL(P)$.

(B) For any finite non-negative integer taken as $ZCL(P)$, determine the corresponding class of P's.

A complete solution of these problems is probably beyond the reach of present-day mathematics. (At least, the author knows of no publication where even an attempt is made to obtain a solution for either one of the general problems indicated.) But more restricted problems in the same field may, in some instances, be solved.

Problem A. When $ZCL(P) > 0$ and n is small, it is easy to construct one or several distinct connected loopless graphs for P, and to show that no other such graph for P exists. Thus, $ZCL(P)$ is determined for the particular P in question. But when P is expressed in terms of variables each of which is allowed to take the value of any finite positive integer, to determine $ZCL(P)$ as a function of these variables may be a matter of considerable difficulty.[2]

Problem B. It is again easy to find by trial and error, partitions which can be proved to correspond to given (small) values of $ZCL(P)$. But the author is not aware of any previous work dealing with problem B in any more general fashion. In the present paper, the classes of partitions for which $ZCL(P) = 0$ and for which $ZCL(P) = 1$ are determined.[3] These results carry with them the determination of the more general class where $ZCL(P) > 1$. When $ZCL(P) = 0$, P will be called *nullgraphic*; when $ZCL(P) = 1$, P will be called *unigraphic*, and the one representative connected loopless graph will be called *unique*; when $ZCL(P) > 1$, P will be called *multigraphic*.

---

[2] As long ago as 1874, Cayley attempted to obtain the value of $ZCL(P)$ for every partition meeting the following condition: $n = 3x + 2$; $p_1 = p_x = 4$; $p_{x+1} = p_{3x+2} = 1$. He was led to consider this particular class of partitions because, for each member of this class, $ZCL(P)$ is the number of structurally isomeric saturated hydrocarbons of the formula C_xH_{2x+2}. These compounds play an important role in structural organic chemistry. Cayley's papers appear in *Brit. Assoc. Adv. Sci. Reports*, p. 275 (1875); *Phil. Mag.* (4), *47*, 444 (1875); (5), *3*, 37 (1877). They are also reprinted in Volume IX of his collected works. Since the publication of these first papers, there have appeared in the chemical literature numerous articles dealing in part with Cayley's problems and in part with problems closely related to his.

[3] Chemists have strong reasons for paying special attention to these two particular values of $ZCL(P)$; moreover, of all possible values of $ZCL(P)$, these are the two for which it is by far the easiest to determine the corresponding classes of partitions.

THEOREM (1.1). *$\Sigma P = 2x$ (where x is any positive integer) is a necessary and sufficient condition for $Z(P) > 0$.*

Where $Z(P) > 0$, let $Ft(P) = \Sigma P/2 - (n - 1)$, and let $Fr(P) = \Sigma P/2 - p_1$.[4]

THEOREM (1.2). *$Ft(P) \geqq 0$ is a necessary and sufficient condition for $ZC(P) > 0$.*

THEOREM (1.3). *$Fr(P) \geqq 0$ is a necessary and sufficient condition for $ZL(P) > 0$.*

THEOREM (1.4). *$Ft(P) \geqq 0 \leqq Fr(P)$ (the conjunction of conditions (1.2) and (1.3)) is a necessary and sufficient condition for $ZCL(P) > 0$. That is to say, if, for any given P, neither the class of connected graphs nor the class of loopless graphs is the null class, then the intersection of these two classes cannot be the null class.*

Hereafter those P's and only those P's which meet condition (1.4) and hence are represented by at least one connected loopless graph will be called Π's.

THEOREM (2). *Any one of the following conditions (2.1)-(2.6″) is sufficient and at least one of them is necessary for $ZCL(\Pi) = 1$.*

UNIQUENESS CONDITIONS.

(2.1)	$Fr(\Pi) = 0$.	*One half of the sum of all the integers in Π is equal to the largest integer in Π.*
(2.2)	$n = 3$.	*The number of integers in Π is 3.*
(2.3)	$p_3 = 1$.	*Π contains at least three integers but not more than two of them are > 1.*
(2.4)	$p_1 = 2$.	*The largest integer in Π is 2.*
(2.5)	$Fr(\Pi) = 1;\ p_2 = p_n$.	*The largest integer in Π is one less than half the sum of all the integers; excepting this largest integer, all the integers in Π are equal.*
(2.6)	$p_1 = p_3;\ p_4 = 1$.	*The three largest integers in Π are all equal to one another; the fourth integer is 1.*
(2.6′)	$n = 4$.	*The number of integers in Π is four.*
(2.6″)	$Ft(\Pi) = 0$.	*The number of integers in Π is $\Sigma\Pi/2 + 1$.*

---

[4] As is more fully explained in Part II of this paper, the designations $Ft(P)$ and $Fr(P)$ arise as follows. If $Ft(P) = 0$, then any connected graph for P is a *tree*; if $Fr(P) = 0$, then any loopless graph for P is a *rosette*.

That is to say, a Π is unigraphic if it meets any one of the above conditions; it is multigraphic if it meets no one of them.[5]

So far, it has been assumed that any two points in C, at least until they have been incorporated in a graph and sometimes even then, are indistinguishable from one another. Consider now in C a set I consisting of $s_i > 1$ points to all of which the same valence p_i is to be assigned. Break I up into disjoint subsets which together exhaust I. These subsets may be made distinguishable from one another by endowing all the points of each particular subset with an intrinsic property such as one particular color, and assigning a different color to each subset. A graph constructed with a set C of points which contains at least one parti-colored set I will be called a *polychromatic* graph to distinguish it from the *monochromatic* graphs hitherto considered.[6]

The identity of two polychromatic graphs is fixed as follows: If G and G' are identical, the corresponding points c_i in G and c_i' in G' must not only be of like valence and similarly attached by edges to themselves and to the other points of G and G' respectively (the purely topological condition which holds for the monochromatic graphs hitherto discussed), but, in addition, the like-valent points c_i' and c_j' in G' must be alike in color if and only if the corresponding like-valent points c_i and c_j in G are alike in color.

THEOREM (3). *The uniqueness conditions for connected loopless graphs must be successively strengthened as the number of colors in these graphs is*

---

[5] The classes of partitions for which $ZCL(P) = 0$ and for which $ZCL(\Pi) = 1$ are both infinite. Moreover the conditions for $ZCL(\Pi) = 1$ are much more complex than those for $ZCL(P) = 0$. It appears probable (though it is not here proved) that the conditions for $ZCL(\Pi) = x$ become increasingly intricate as x increases. Probably, for every positive value of x, the class of corresponding Π's is infinite. But the author knows of no proof that such is the fact; indeed, there is no proof that, for at least one value of x, the class of corresponding partitions may not be the null class.

For structural organic chemistry, no partition which is not a Π has any significance, and all Π's are subject to the mathematically arbitrary but chemically important limitation $p_1 < 5$. When this limitation is imposed, the class of unigraphic partitions shrinks to two infinite sequences (where $p_1 = 2$) plus about thirty individually defined members. Since no member of either of the two infinite sequences has any chemical importance when $n > 4$, the entire number of unigraphic partitions which might possibly be of chemical significance shrinks to less than 35. To examine the chemical representatives (if any exist) of these partitions in order to see whether they possess the proper chemical requirements, is a brief task.

[6] So far as the number of distinct graphs for a given P is concerned, polychromism among points of unlike valence is a vacuous phenemenon. This assertion follows from the fact that, in matching the points of two graphs G and G' to see whether these graphs are or are not distinct, no two points of unlike valence can ever correspond to one another, whether they are of the same color or not.

increased from one to four; a unique connected loopless graph with more than three colors is a rosette.

Of the theorems so far given, (1.1)-(1.4) are proved in part II, (2) is proved in part III, and (3) is proved in part IV of the present paper. In each instance, the proofs cannot well be given until certain methods of constructing and classifying graphs have been described. The necessary definitions and descriptions therefore precede the actual proofs in each of the three following parts.

Part II. Proof of theorems (1.1)-(1.4).

Symbolism and terminology. If, in the graph G, any part H is a connected graph, and if (excepting H itself) no part of G which includes H is a connected graph, then H is a *component* of G. The number of distinct components which make up the whole of G is $\nu(G)$, and where G is a finite graph, $\nu(G)$ is a positive finite integer. The expression $\nu(G) = 1$ means that G itself is connected. If and only if, in G, at least one edge is terminated by both points c_i and c_j, then these two points form in G an *attached* pair. This term 'attached' is far stronger than the term 'connected' which, when applied to a pair of points, refers to any two points in one and the same component of G.

The set of all those edges which attach c_i to c_j is called the *bond* (i, j). The number of edges in the bond (i, j) is called the *breadth* of that bond and is indicated by the symbol $b(i, j)$. When $b(i, j) = 1$, the bond is *single*; when $b(i, j) > 1$, the bond is *multiple*. Two points c_i and c_j together with the q edges which form the bond between them may be indicated in a diagram by $c_i \overset{q}{\text{——}} c_j$ and symbolically by (c_i, q, c_j). This symbol implies and is implied by the equation $b(i, j) = q$. The scheme just described may be extended to cover the case (c_i, q, c_i) or $b(i, i) = q$ where q is the number of loop-edges dependent from c_i. Any graph for P is uniquely determined if the value is given for every $b(i, j)$ where $i = 1, \cdots, n$ and $j = 1, \cdots, n$. In all cases, $b(i, j) = b(j, i)$, and $p_a = \sum_{i=n}^{i=1} b(a, i) + b(a, a)$. Where, for a particular pair of points (c_i, c_j), $b(i, j) = 0$, that fact will usually be indicated by omitting the equation and the corresponding symbol $(c_i, 0, c_j)$; but occasionally it is convenient to use such an equation or such a symbol explicitly. Particularly, when one equation or one symbol is used to describe the bond (i, j) in various different graphs, then $b(i, j)$ may be expressed in terms of variables. For example, in the equation $b(i, j) = q - x$ and in the symbol

$(c_i, q - x, c_j)$, let q be a constant and x a variable. Both are positive integers. The scheme just outlined permits $x = q$ as a special case, but under no circumstances may $b(i, j)$ be negative. This same symbolic scheme may be still further extended to apply to subgraphs containing more than two points. The meanings of the symbols $(c_i, b(i, j), c_j, b(j, k), c_k)$ and $(c_i, b(i, j), c_j, \cdots, c_u, b(u, v), c_v)$ should be self-evident. Such a symbol corresponds to a set of equations, not to any single equation.

If $b(i, j) = p_i$, then c_i is a *blind point.* If, in addition, $b(i, j) = 1$, then c_i is an *end point,* and the single edge attaching c_i to c_j is an *end edge.* If, in any connected loopless graph, the number of blind points exceeds $n - 2$, the graph is a *rosette.* If all the blind points in a rosette are like-valent, the rosette is *regular*; otherwise it is *irregular.* A regular rosette in which all the blind points are end points is a *star.*

Rosettes are an important and striking class of connected loopless graphs. They are characterized by the property $Fr(\Pi) = 0$, and have the following other properties:

(1) According to condition (1.4), rosettes border the field of connected loopless graphs on one side.

(2) According to condition (2.1), every rosette is unique.

(3) If any Π is represented by a rosette, it cannot be represented by any other loopless graph, connected or disconnected.

If c_1 terminates every non-loop edge in a graph and is moreover the only loop-point in that graph, then the graph will be called a *loop-rosette.* This definition includes as special cases those connected graphs in which the number of non-loop edges is 0.

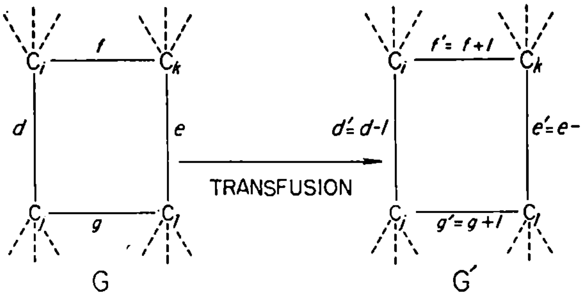
Any edge which forms part of a cycle in G will be said to be *cyclic* in G. If $b(i, j) = q > 1$, then all the edges in the bond (i, j) are cyclic in G, and at most $q - 2$ of them may always be expunged without destroying the cyclic character of the remaining edges. If $q - 1$ of the edges are expunged, the one remaining edge is non-cyclic *unless,* in G, c_i and c_j are both members of some one and the same cycle of order > 2. A connected graph which has no cyclic edge is a *tree.*

Since every graph uniquely defines its partition, the expressions $Ft(G)$ and $Fr(G)$ may legitimately be used. Furthermore, if G and G' both represent the same P, then $Ft(G) = Ft(G')$ and $Fr(G) = Fr(G')$. For a tree, $Ft(G) = Ft(P) = \Sigma P/2 - (n - 1) = 0$; for a rosette, $Fr(G) = Fr(P) = \Sigma P/2 - p_1 = 0$.

Transfusion. Consider the graph G containing the points c_i, c_j, c_k and c_l. Let $b(i,j) = d$, $b(k,l) = e$, $b(i,k) = f$, $b(j,l) = g$. Decrease by one the breadths of the bonds (i,j) and (k,l) by expunging one edge from each of these bonds; increase by one the breadths of the bonds (i,k) and (j,l) by adding one edge to each of these bonds. By this process there is formed the new graph G' where $d' = d - 1$, $e' = e - 1$, $f' = f + 1$, $g' = g + 1$. It is *essential* for the carrying out of this process that, in G, $d > 0 < e$—i. e., that in G the pairs of points (c_i, c_j) and (c_k, c_l) be attached pairs. It is *inevitable* that, after the carrying out of the process, $f' > 0 < g'$—i. e., that in G' the pairs of points (c_i, c_k) and (c_j, c_l) be attached pairs. Hence the process just described is called the *transfusion* of the bond (i,j) with the bond (k,l).[7] As for the bonds (i,k) and (j,l) in G as well as the bonds (i,j) and (k,l) in G', their breadths may be equal to or > 0, depending on the particular G on which the transfusion is performed. (Any bonds in G and G' other than the four here mentioned are common to both of these graphs; they are not involved in the transfusion in question and are irrelevant to that process.) Symbolically the transfusion just described may be expressed by the following equations:

$$b'(i,j) = b(i,j) - 1, \qquad b'(i,k) = b(i,k) + 1,$$
$$b'(k,l) = b(k,l) - 1, \qquad b'(j,l) = b(j,l) + 1.$$

Diagrammatically expressed, it is as follows:



In a transfusion, no one of the points involved undergoes any change in valence, and hence, *P is invariant under transfusion.* In the three possible cases where a transfusion decreases, leaves invariant or increases the number of components, the transfusion will be called respectively *connective, neutral* or *disconnective.* Various special cases demand attention.

---

[7] Note that when a transfusion is employed in a construction, the pairs of points (i,j) and (k,l) are to be regarded as ordered pairs. The instruction "Transfuse (i,j) with (k,l)" means that the new edges to be introduced are (i,k) and (j,l). On the other hand, the instruction "Transfuse (i,j) with (l,k)" means that the new edges to be introduced are (i,l) and (j,k). The edges to be expunged are (i,j) and (k,l) in both instances.

I. The two bonds (i, j) and (k, l) belong to different components of G. Here a transfusion may be connective or neutral but cannot be disconnective. *The transfusion is connective if and only if at least one of the two expunged edges is cyclic in G.* Where this condition is met $\nu(G') = \nu(G) - 1$.

II. The two bonds (i, j) and (k, l) belong to one and the same component H of G. Here a transfusion may be neutral or disconnective but cannot be connective. The transfusion is disconnective if and only if the two expunged edges form in H two single bonds which constitute the sole connection between that part of H which contains c_i and c_k and that part of H which contains c_j and c_l. Where this condition is met, $\nu(G') = \nu(G) + 1$. For the present discussion, the most interesting transfusions within one component of G are those where at least one of the expunged edges is a loop-edge. Any such process will hereafter be called a *loop-transfusion.*

III. For purposes of transfusion, a loop-point may be regarded as self-attached. Two varieties of loop-transfusion may be recognized:

(a) The two expunged edges are loop-edges dependent respectively from the two distinct loop-points c_k and c_m.

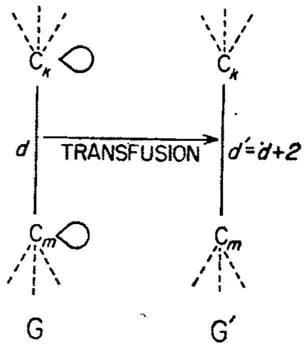
$$b'(m, m) = b(m, m) - 1, \quad b'(k, k) = b(k, k) - 1, \quad b'(k, m) = b(k, m) + 2.$$

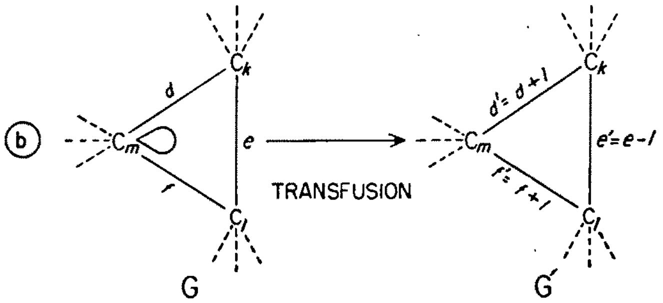
(b) One of the two expunged edges is a loop-edge dependent from the loop-point c_m; the other is an edge in the bond (k, l).

$$b'(m, m) = b(m, m) - 1, \qquad b'(k, m) = b(k, m) + 1,$$
$$b'(k, l) = b(k, l) - 1, \qquad b'(l, m) = b(l, m) + 1.$$

When diagrammatically expressed, these two processes are as follows:

ⓐ





In diagram (a), d may be equal to or > 0, but d' is necessarily > 0; in diagram (b), d and/or f and/or e' may be equal to or > 0, but e, d', and f' are necessarily > 0.

Since a loop-edge is always cyclic, and since it can belong to no cycle of order > 1, a loop-transfusion is connective if the two expunged edges belong to different components of G; it is neutral if they belong to the same component of G. *Under no circumstances can a loop-transfusion be disconnective.* Process (a) can be carried out if and only if G contains at least two distinct loop-points; it always decreases the number of loops by two. Process (b) always decreases the number of loops by one; but this process is subject to an important limitation. It can be carried out if and only if c_k and c_l constitute an attached pair each member of which is distinct from loop-point c_m. This last condition can always be met if G has at least one component H_i which includes more than one point, and which is distinct from the component H_j containing c_m. But if every non-loop edge in G terminates in c_m, process (b) fails. Hence, if and only if,

(1) c_m is the only loop-point in G, and

(2) every non-loop edge in G terminates in c_m,

that is, if and only if G is a loop-rosette, then processes (a) and (b) both fail.

The important conclusions to be drawn from the above considerations are:

(1) If and only if G is connected or G contains no cyclic edge, a connective transfusion in G is impossible.

(2) If and only if G is loopless or a loop-rosette, a loop-transfusion in G is impossible.

(3) The number of components in G cannot be increased by sucessive loop-transfusions.

(4) The number of loops in G cannot be increased by successive connective transfusions.

Although the object of the procedures hereafter described is to construct loopless connected graphs, it is sometimes convenient, as an intermediate step in such a construction, to reverse a loop-transfusion of type (b). Such a procedure will be called a *reversion,* and when ν is invariant under a particular reversion, that reversion will be called a *neutral reversion.* A neutral reversion can always be performed on a graph G provided that G contains a subgraph defined by $b(m,k) > 1$, $b(m,l) > 0$. Where this condition is met, it is possible, by a neutral reversion, to form a graph G' which contains the subgraph defined by $b(m,m) = 1$, $b(m,k) > 0$, $b(k,l) > 0$. Here $\nu(G) = \nu(G')$.

Proof of Theorem (1.1). Theorem (1.1) states that $\Sigma P = 2x$ (where x is any positive integer) is a necessary and sufficient condition for $Z(P) > 0$. The necessity of condition (1.1) follows directly from the fact that every edge has just two ends. The sufficiency of this same condition is shown by the following construction.

Divide P into two partitions Pe and Po, the first consisting of all the even integers $e_1, e_2, \cdots, e_k$ in P, the second consisting of all the odd integers $o_1, o_2, \cdots, o_{n-k}$ in P. According to condition (1.1), $(n-k)$ is even. Make each point ce_i the loop-point of $e_i/2$ loops; make each point co_i the loop-point of $(o_i - 1)/2$ loops; then attach each point co_{2i} to the point co_{2i-1} by a single edge. By this procedure, a graph for P is formed and the sufficiency of condition (1.1) for $Z(P) > 0$ is proved.

A consequence of Theorem (1.1) is the following

LEMMA. *If all the integers of* P_1 *are divided into* $y-1$ *disjoint sets which constitute respectively the partitions* $P_2, \cdots, P_y$, *and if all but one of these partitions* $P_1, \cdots, P_y$ *are represented by graphs, then there must always be at least one graph for the remaining y-th partition. This rule is particularly useful where each of the disjoint sets of integers is composed of integers which are consecutive in* P_1.

In all the following discussion, it will be assumed that ΣP is an even positive integer, and hence that at least one graph for P exists.

Proof of Theorems (1.2)-(1.4). Theorems (1.2)-(1.4) are existence theorems. Each one states that if and only if P meets a certain condition, at least one graph for P has the corresponding property. They do *not* imply that if and only if P meets the condition in question every graph for P has that property.[8] Consequently, the three following proofs are similar in outline. A general method is given for constructing for P a graph with the desired property. It is then proved that the given method fails if and only if P does not meet the corresponding condition.

(A) Theorem (1.2) states that $\Sigma P/2 - (n-1) = Ft(P) \overset{=}{>} 0$ is a necessary and stufficient condition for $ZC(P) > 0$. For the proof of this theorem, further consideration of the function $Ft(P)$ is necessary.

$Ft(P)$ is closely related to but not identical with the well-known graph function μ, called the cyclomatic number of a graph. The function μ is usually defined by $\mu = \alpha_1 - \alpha_0 + \nu$ where α_1 is the number of edges, α_0 is the number of points (vertices), and ν is the number of components of the graph. In the notation here used, $\alpha_1 = \Sigma P/2$ and $\alpha_0 = n$; hence $\mu = \Sigma P/2 - n + \nu$. But ν (although it is a graph function) is (usually) not uniquely determined by P. Moreover, $\mu(G)$ is specially designed so that $\mu(G) < 0$ is an impossibility, whereas $Ft(P)$ is useful chiefly because it sometimes has a negative value. The exact relation between $Ft(G)$ and $\mu(G)$ is as follows:

$$Ft(G) = \Sigma P/2 - n + 1, \qquad \mu(G) = \Sigma P/2 - n + \nu.$$

$$\text{Therefore } Ft(G) = \mu(G) - (\nu - 1).$$

That is to say: $Ft(G) = \mu(G)$ when G is connected, $(\nu = 1)$.

$Ft(G) < \mu(G)$ when G is not connected, $(\nu > 1)$.

It is well known that, where G consists of ν components $H_1, H_2, \cdots, H_\nu$,

(a) if $\nu = 1$ and $\mu(G) = 0$, then G is a tree;

(b) $\mu(G) = \mu(H_1) + \mu(H_2) + \cdots + \mu(H_\nu)$.

Construct for P a graph G_0 which may or may not contains loops and may or may not be connected. Next perform on G_0 a connective transfusion to form G_1. This latter step must always be possible except in the two following instances:

(1) $\nu_0 = 1$. Here $Ft(P) = Ft(G_0) = \mu(G_0) - (\nu_0 - 1)$ reduces to $Ft(P) = \mu(G_0)$; and therefore $Ft(P) \overset{=}{>} 0$.

---

[8] Evidently, in the highly special cases where $Z(P) = 1$, an existence proof for a particular kind of graph for P becomes a universal proof. Hence the following table.

p_1	n	$\Sigma P/2$	$Ft(P)$	$Fr(P)$	The only possible graph is:
4	1	2	2	-2	connected but not loopless.
1	4	2	-1	1	loopless but not connected.
1	2	1	0	0	loopless and connected.

(2) $\nu_0 > 1$, but G_0 contains no edge cyclic in G_0. Here $H_1, H_2, \cdots, H_\nu$ (the components of G_0) must all be trees. Each $\mu(H_i) = 0$ and hence $\mu(G_0) = 0$. Here $Ft(P) = Ft(G_0) = \mu G_0 - (\nu_0 - 1)$ reduces to $Ft(P) = 1 - \nu_0$; and, since $\nu_0 > 1$, $Ft(P) < 0$. Here $1 - Ft(P) = \nu_0$, the number of trees in G_0.

If $\nu_0 > 1$, and G_0 contains at least one edge cyclic in G_0, then the graph G_1 formed by a connective transfusion performed on G_0 contains $\nu_0 - 1$ components. Exactly the same argument used for G_0 may now be used for G_1. This argument shows that G_1 is connected $(Ft(P) \gtreqless 0)$, or that G_1 consists of $\nu_1 = (\nu_0 - 1) > 1$ trees $(Ft(P) < 0)$, or that a connective transfusion may be performed on G_1, by which transfusion G_1 is converted into G_2 with $\nu_0 - 2$ components. Successive connective transfusions performed on G_0, G_1, etc., must therefore lead to one of two results:

(a) If $\mu(G_0) \gtreqless (\nu_0 - 1)$, then $Ft(P) \gtreqless 0$ and $G_{\nu_{0}-1}$ is a connected graph.

(b) If $\mu(G_0) < (\nu_0 - 1)$, then $Ft(P) < 0$ and $G_{\mu(G_0)}$ is a set of $1 - Ft(P)$ trees.

It is thus proved that $Ft(P) = 0$ is a necessary and sufficient condition for $ZC(P) > 0$.

(B) Theorem (1. 3) states that $\Sigma P/2 - p_1 = Fr(P) \gtreqless 0$ is a necessary and sufficient condition for $ZL(P) > 0$.

Construct for P a graph G_0 which may or may not contain loops and may or may not be connected. Next perform on G_0 a loop-transfusion to form G_1. This latter step must always be possible except in the two following instances:

(1) G_0 is loopless. Here $Fr(G_0) = Fr(P)$ is the number of the edges in G_0 which are *not* terminated by c_1. Evidently, this number cannot be negative, and hence $Fr(P) \gtreqless 0$.

(2) G_0 is a loop-rosette. Here $p_1 = p_2 + p_3 + \cdots + p_n + 2s$ where $s\ (> 0)$ is the number of loop-edges dependent from c_1. Hence

$$2p_1 > (p_1 + p_2 + \cdots + p_n) = \Sigma P;\ p_1 > \Sigma P/2;$$
$$\text{and } Fr(P) = (\Sigma P/2 - p_1) < 0.$$

If G_0 is neither loopless nor a loop-rosette, then the number of loops in G_1 must be less than the number of loops in G_0. Exactly the same argument used for G_0 may now be used for G_1. This argument shows that G_1 is loopless ($Fr(P) \eqslantgtr 0$) or that G_1 is a loop-rosette ($Fr(P) < 0$) or that a loop-transfusion may be performed on G_1, by which transfusion G_1 is converted into G_2 which contains less loops than are contained in G_1. Successive loop-transfusions performed in G_0, G_1, etc., must therefore lead to one of two results:

(a) G_i is loopless, and here $Fr(P) \eqslantgtr 0$.

(b) G_i is a loop-rosette, and here $Fr(P) < 0$.

That is to say, the only condition which forbids the construction of a loopless graph for P is the only condition which permits $Fr(P)$ to have a negative value. Thus it is proved that $Fr(P) \eqslantgtr 0$ is a necessary and sufficient condition for $ZL(P) > 0$.

(C) Theorem (1. 4) states that $Ft(P) \eqslantgtr 0 \eqslantless Fr(P)$ (the conjunction of conditions (1. 2) and (1. 3)) is a necessary and sufficient condition for $ZCL(P) > 0$. Assume that $Ft(P) \eqslantgtr 0 \eqslantless Fr(P)$. Construct a graph G for P. If G is not connected, convert it by successive connective transfusions into a connected graph G' for P. This procedure must be possible since it has been proved to fail if and only if $Ft(P) < 0$. Next, if G' is not loopless, convert it by successive loop-transfusions into a loopless graph G'' for P. This procedure must be possible since it has been proved to fail if and only if $Fr(P) < 0$. The graph G'' thus finally obtained must be connected and loopless because G' is connected and (as has already been shown) no loop-transfusion can be disconnective. It would evidently be impossible to carry out the full procedure just described if either $Ft(P)$ or $Fr(P)$ were negative. Hence $Ft(P) \eqslantgtr 0 \eqslantless Fr(P)$ is a necessary and sufficient condition for $ZCL(P) > 0$.[9]

So far it has been shown that any graph for P may be converted by transfusions into

(1) a connected loopless graph: $Ft(P) \eqslantgtr 0$; $Fr(P) \eqslantgtr 0$; or

(2) a set of ν (> 1) trees—loopless but not connected: $Ft(P) < 0$; $Fr(P) \eqslantgtr 0$; or

(3) a loop-rosette—connected but not loopless: $Ft(P) \eqslantgtr 0$; $Fr(P) < 0$.

---

[9] It is of course possible to reverse the order of the transfusions used in the above proof. That is, G may first be converted by loop-transfusions into a loopless graph, and then (since no connective transfusion can increase the number of loops in a graph) by connective transfusions into a connected loopless graph. The result is the same, whichever method is adopted.

The only other logical possibility is $Ft(P) < 0$; $Fr(P) < 0$. But here $Ft(P) + Fr(P) < 0$; that is,

$$\Sigma(P)/2 - (n-1) + \Sigma(P)/2 - p_1 = \Sigma(P) - (n-1) - p_1 < 0;$$
$$p_2 + p_3 + \cdots + p_n < n-1;$$

and this inequality is impossible if $p_2, p_3, \cdots, p_n$ are all positive integers. That is to say, there is no P which is represented only by disconnected graphs with loops.[10]

Part III. Proof of theorem (2).

(A) The sufficiency of each one of the conditions (2.1)-(2.6″) for $ZCL(\Pi) = 1$.

Each of the uniqueness conditions (2.1)-(2.6″) (p. 666) is independent of all the others. This assertion is proved by the following Π's each one of which fulfills that condition and only that condition to which it corresponds in the following list.

Condition.	Π.	Condition.	Π.
(2.1)	3, 3	(2.5)	4, 2, 2, 2
(2.2)	4, 4, 4	(2.6′)	3, 3, 3, 1
(2.3)	3, 3, 1, 1	(2.6″)	3, 3, 3, 1, 1, 1, 1, 1
(2.4)	2, 2, 2, 2		

But, as thus stated, the uniqueness conditions partially overlap (e.g. the partition 2, 1, 1 fulfills each of the conditions (2.1)-(2.4)), and hence they cannot without some redundancy be proved to be individually sufficient. In such a demonstration it is better to be guided by the following rubric in which the same uniqueness conditions are restated in a form which prevents overlapping.

RUBRIC.

$Fr(\Pi) = 0$.	Class I.
$Ft(\Pi) > 0$.	
$n = 3$.	Class II.
$n > 3$.	

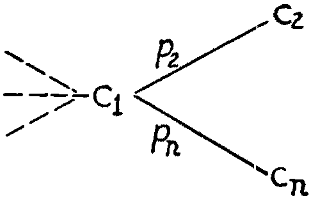
---

[10] One inference to be drawn from the fact that $Ft(P) + Fr(P) \geqq 0$ is that the value of $Ft(P)$ sets a lower limit to the value of $Fr(P)$ and vice versa. There is, in fact, a quasi-dualism between $Ft(P)$ and $Fr(P)$ which is associated with the well-known dualism among partitions. If P_d is the partition dual to P, then $\Sigma P = \Sigma P_d$; $n = p_{1_d}$; $p_1 = n_d$; and hence $Ft(P) + Fr(P) = Ft(P_d) + Fr(P_d)$; $Ft(P) = Fr(P_d) + 1$; $Fr(P) = Ft(P_d) - 1$. This quasi-dualism does not, however, lead to any precise dualism between rosettes and trees.

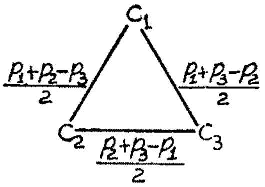
- $p_3 = 1$. Class III.
- $p_3 > 1$.
  - $p_1 = 2$. Class IV.
  - $p_1 > 2$.
    - $Fr(\Pi) = 1$; $p_2 = p_n$. Class V.
    - $Fr(\Pi) > 1$; $p_1 = p_3$; $p_4 = 1$.
      - $n = 4$. Class VI.
      - $Ft(\Pi) = 0$. Class VII.

The general graphs for the Π's of classes I-VII are given below. All the Π's of class IV fall into two disjoint subclasses: IVa, where $p_n = 2$, and IVb, where $p_n = 1$.

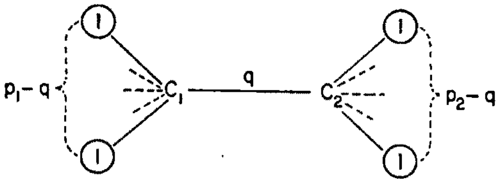
Class I. $b(1, i) = p_i$;
G is a rosette. $i = 2, \cdots, n$.



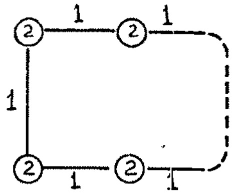
Class II. $b(1, 2) = (p_1 + p_2 - p_3)/2$;
$b(2, 3) = (p_2 + p_3 - p_1)/2$;
$b(3, 1) = (p_1 + p_3 - p_2)/2$.



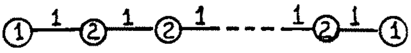
Class III. $b(1, 2) = q$; $0 < q < p_2$;
$b(1, i) = 1$; $i = 3, \cdots, (p_1 - q + 2)$;
$b(2, j) = 1$; $j = (p_1 - q + 3), \cdots, n$.



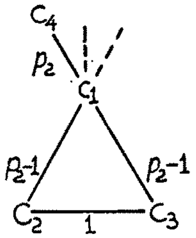
Class IVa. $b(i, i + 1) = 1$; $i = 1, \cdots, (n - 1)$;
$b(n, 1) = 1$.



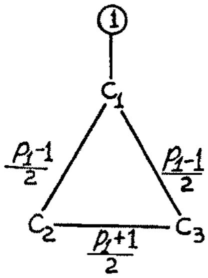
Class IVb. $b(1, n) = 1$;
G is a tree. $b(i, i+1) = 1; i = 1, \cdots, (n-2)$.



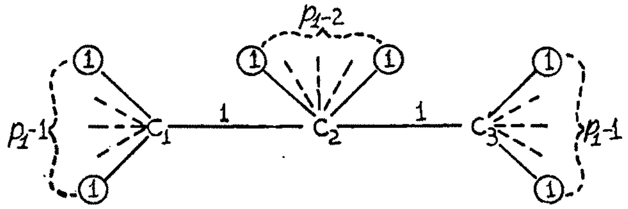
Class V. $b(1,2) = b(1,3) = (p_2 - 1)$;
$b(2,3) = 1$;
$b(1,i) = p_2; i = 4, \cdots, n$.



Class VI. $b(1,2) = b(1,3) = (p_1 - 1)/2$;
$b(2,3) = (p_1 + 1)/2$;
$b(1,4) = 1$.



Class VII. $b(1,2) = b(2,3) = 1$.
G is a tree.
$b(1,i) = 1; i = 4, \cdots, (p_1 + 2)$;
$b(2,j) = 1; j = (p_1 + 3), \cdots, 2p_1$;
$b(3,k) = 1; k = (2p_1 + 1), \cdots, n$.



In every case except class V, the uniqueness of the general graph given for the Π's in question is obvious. The uniqueness of the general graph given for any Π of class V is made evident by the following considerations. Every graph for a Π of class V may be constructed as follows. Where $Fr(\Pi) = 0$, $n > 3$ and $p_2 = p_n > 1$, construct a graph for Π. This graph is a rosette and therefore unique (Class I); it is moreover a regular rosette. In this rosette, expunge two edges so as to decrease $b(1,i)$ and $b(1,j)$ each from p_2

to $(p_2 - 1)$; insert a new edge so as to increase $b(i, j)$ from 0 to 1. A graph for a Π off class V is thus formed. Since all the pairs of points (c_i, c_j) which can be chosen from $c_2, c_3, \cdots, c_n$ are equivalent in the regular rosette, all possible ways of carrying out the indicated construction are equivalent, and the general graph given for any Π of class V is unique.

The sufficiency of each one of the conditions (2.1)-(2.6″) for $ZCL(\Pi) = 1$ is thus demonstrated.

(B) The necessity of at least one of the conditions (2.1)-(2.6″) for $ZCL(\Pi) = 1$.

The full proof that at least one of the conditions (2.1)-(2.6″) is necessary for $ZCL(\Pi) = 1$ is extremely laborious. For this purpose, no method better than the following has so far been found. The Π's are exhaustively classified into disjoint classes. For each of these classes, a method is given for constructing two distinct connected loopless graphs for every member of the class which is in no one of Classes I-VII. The success of this scheme depends on finding a suitable way of defining the disjoint classes of Π's. For, if the number of these classes is too large, the proofs become insufferably repetitious; and if, on the other hand, the number of these classes is too small, then, in some instances, the general method of construction applicable to all the Π's in the class under consideration becomes so intricate that it cannot readily be grasped.

A general principle for classifying Π's is therefore next given. Then follow the definitions of certain concepts which serve to abbreviate the arguments in the ensuing proof.

Classification of Π's; complete expansion sequences. Consider a sequence of connected loopless graphs G_0, G_1—in which each G_i is formed by deleting one cyclic edge in G_{i-1}. Since each G_i uniquely defines its Π_i, the corresponding sequence of Π's may also be considered. In such a *deletion sequence*, G_i does not (usually) uniquely determine either G_{i+1} or G_{i-1}; and Π_i, although it puts strong limitations on both Π_{i+1} and Π_{i-1}, does not (usually) uniquely determine either of these partitions. The maximum number of successors to Π_i in a deletion sequence is $Ft(\Pi_i)$; but no property of Π_i permits any maximum to be set on the number of its predecessors in the sequence.

It is, however possible to devise a sequence of graphs similar in some respects to a deletion sequence, but possessing the following useful properties:

(a) Π_i uniquely determines both Π_{i+1} and Π_{i-1}, and hence every Π in the sequence.

(b) Both the maximum number of successors and the maximum number of predecessors of Π_i in the sequence are uniquely determined by Π_i.

For the construction of such a sequence, consider a connective transfusion between the connected loopless graph G_i and the unique graph $(c_y, 1, c_z)$ (for the partition 1, 1) to form the graph G_{i+1}. This particular kind of transfusion, where the edges expunged are any one cyclic edge in G_i and the only edge in $(c_y, 1, c_z)$, will be called a *one stage expansion* of G_i. Here evidently $Ft(G_{i+1}) = Ft(G_i) - 1$, and so, in an *expansion sequence* of connected loopless graphs where G_{i+1} is formed by the one stage expansion of G_i, the maximum number of successors or *expansion products* of G_i is $Ft(G_i)$. In such a sequence, G_i does not (usually) uniquely determine either G_{i+1} or G_{i-1}; but Π_i does uniquely determine both Π_{i+1} and Π_{i-1}, for Π_{i+1} is merely Π_i with two unit integers adjoined, and Π_{i-1} is merely Π_i with two unit integers omitted.

In the process of *one-stage contraction* (which is the inverse of one-stage expansion), two end points attached to two distinct points c_j and c_k of G_i (together with their end-edges) are expunged, and a new edge attaching c_j to c_k is inserted to form G_{i+1}. No sequence of one-stage contractions can be continued indefinitely, for one obvious limitation on contraction is the following. Let $Fu(\Pi)$ be the number of unit integers in Π; and let $Fu'(\Pi)$ be one half of the largest even number $\leqq Fu(\Pi)$. An algebraic expression for $Fu'(\Pi)$ is

$$Fu'(\Pi) = \tfrac{1}{4}\{2Fu(\Pi) - 1 + (1)^{Fu(\Pi)}\}.$$

The maximum number of successors of Π_i in a *contraction sequence* cannot exceed $Fu'(\Pi_i)$. But there is another limitation on the number of such succesors. The integer p_1 in Π_i is identical with the integer p_1 in Π_{i+1}; but $\Sigma\Pi_i/2 = \Sigma\Pi_{i+1}/2 + 1$. Hence $Fr(\Pi_{i+1}) = Fr(\Pi_i) - 1$. It may be that, after $Fr(\Pi_i)$ successive contractions, of which the first is performed on Π_i, the stage is reached where $Fr(\Pi_{i+Fr(\Pi_i)}) = 0$. In this case $G_{i+Fr(\Pi_i)}$ is a rosette; consequently, even if $Fu'(\Pi_{i+Fr(\Pi_i)}) > 0$, no further contraction is possible without the formation of a forbidden loop. Since the successors of Π_i in a contraction sequence are respectively the predecessors of the same Π_i in an expansion sequence, the maximum number of such predecessors is the lesser of the two functions $Fr(\Pi_i)$ and $Fu'(\Pi_i)$. (Where $Fr(\Pi_i) = Fu'(\Pi_i)$, the maximum number of predecessors is given by either of these two functions. Here the first graph in the expansion sequence is a rosette with < 2 end points.) If the lesser of $Fr(\Pi)$ and $Fu'(\Pi)$ is called $Fm(\Pi)$, then an algebraic expression for that function is

$$Fm(\Pi) = \tfrac{1}{2}\{Fr(\Pi) + Fu'(\Pi) - |(Fr(\Pi) - Fu'(\Pi)|\}.$$

A Π for which $Fm(\Pi) = 0$ will be called an *initial* Π and will be characterized by the symbol Π_0. A Π_0 is easily recognized, since either $Fr(\Pi_0) = 0$ or $Fu(\Pi_0) < 2$. There are of course Π_0's which meet both of these conditions, and so a connected loopless graph for a Π_0 is a graph which is a rosette or a graph containing < 2 end-points or both.[11]

A Π for which $Ft(\Pi) = 0$ will be called *final*. It is easy to recognize such a final Π by its defining condition. The connected loopless graphs for final Π's are trees, and a connected loopless graph for a unigraphic final Π is a unique tree. Necessary and sufficient conditions for uniqueness among trees (over and above the general condition $Ft(\Pi) = 0$ which holds for all trees) are (a) $p_1 = p_4 = 2$; or (b) $p_1 = p_3 > p_4 = 1$; or (c) $p_2 > p_3 = 1$; or (d) $p_2 = 1$. These conditions are included in the conditions (2. 1), $\cdots$, (2. 6'').

Let any Π which is neither initial nor final be called *intermediate*. In an intermediate Π,

$$\text{(a)} \quad Ft(\Pi) > 0 < Fr(\Pi); \qquad \text{(b)} \quad p_{n-1} = 1.$$

An expansion sequence which begins with an initial and ends with a final Π will be called a *complete expansion sequence*. The number of members of a complete expansion sequence is $Fm(\Pi_i) + Ft(\Pi_i) + 1$ where Π_i is any member of the sequence. Where Π_i is a Π_0, the number of members of the complete sequence is $Ft(\Pi_0) + 1 = \mu(G_0) + 1$.

Expansion sequences lead to a very useful scheme for the classification of all Π's. Construct the disjoint classes (which together exhaust all Π's) in such a way that each class contains only complete expansion sequences. By means of this principle, the problem of classifying all Π's is reduced to the problem of classifying all initial (or all final) Π's—an incomparably easier task.

Inspection of Classes I-VII (pp. 678-679) shows that:

The unigraphic Π's of Class I are initial (sometimes also final).

The unigraphic Π's of Classes II, IVa, V and VI are initial.

---

[11] In constructing the successive graphs of a contraction sequence, some difficulty may be caused by the fact that occasionally, although Π_i is not initial, all of the end-points in G_i are attached to one point c_j. If this situation occurs, proceed as follows. Replace two of the end-points and their end-edges in G_i by a loop-edge dependent from c_j; then remove the loop-edge by a loop-transfusion. Such a transfusion must always be possible wherever G_i (as in the instance under consideration) is not a rosette.

The unigraphic Π's of Class III are intermediate or final.

The unigraphic Π's of Classes IVb and VII are final.

The same inspection reveals that every final Π (classes I, III, IVb or VII) belongs to a complete expansion sequence in which Π_0 is unigraphic and of Class I (where $p_2 = 1$, or $p_2 = p_1 > 1$, or $p_3 = 1$) or Class II (where $p_1 = p_3$) or Class IVa or Class VI. A Π_0 which meets any one of these conditions will be called a $\underline{\Pi}_0$.

Since necessary and sufficient conditions for uniqueness among trees are almost self-evident and have long been known, it is safe to assume that there are no unigraphic final Π's besides those in the classes listed above. Hence the

LEMMA. *A necessary and sufficient condition for a unigraphic final* Π *is that this final* Π *belong to a complete expansion sequence in which* Π_0 *is a* $\underline{\Pi}_0$.

Preservand subgraphs. Pick from any connected loopless graph G a connected subgraph V which is not the whole of G. One of two alternatives must be true:

$$\text{(a)} \quad Ft(G) = Ft(V) \quad \text{or} \qquad \text{(b)} \quad Ft(G) > Ft(V).$$

In the latter case, G may be converted by successive expansions into a connected graph G_i, where $Ft(G_i) = Ft(V)$, and G_i contains V intact. The method of expanding G to G_i is to expunge from G and its successive expansion products only cyclic edges which are not in V. The subgraph V thus preserved intact in the expansion sequence (as far as the stage where $Ft(G_i) = Ft(V)$) will be called a *preservand subgraph* of G.

Application of the idea of a preservand subgraph leads at once to the following

LEMMA. *If*

(1) G_i *and* G_i' *are distinct representatives of one and the same* Π_i *where* $Ft(\Pi_i) > 0$, *and*

(2) G_i *contains a preservand subgraph* V *which distinguishes* G_i *from* G_i' *and*

(3) V *contains no point which is an end-point in* G_i, *and*

(4) $Ft(V) < 2$,

then no explanation product of Π_i *is unigraphic unless* Π_i *belongs to a complete expansion sequence beginning with a* $\underline{\Pi}_0$; *and if* Π_i *does belong to a complete expansion sequence beginning with a* $\underline{\Pi}_0$, *then its only unigraphic expansion product is the final* Π.

A preservand subgraph V which meets the conditions just stated will be called a $\bar{V}$. Naturally, the above lemma is particularly useful where Π_i is a Π_0. The presence of a $\bar{V}$ in one of two distinct representatives (G_0 and G_0') of Π_0 implies that no expansion product of Π_0 is unigraphic.

Additional terminology.

(1) *Polarity.* Where $b(i, j) > 0$, if $p_i = p_j$, the bond (i, j) as well as each of the edges which together form that bond are ***homopolar***; where $p_i \neq p_j$, the bond (i, j) and the edges which compose it are ***heteropolar***.

(2) *Alternating and uniform partitions.* There are certain classes of partitions which, so far as the construction of one representative graph is concerned, lend themselves to particularly simple treatment. These are the *alternating* and *uniform* partitions hereafter described and denoted by the respective symbols Pa and Pu. Almost all alternating and uniform partitions are multigraphic, but each one of them has among its representative graphs one uniquely definable graph hereafter called respectively the *alternating* or the *uniform graph*. These are denoted by the respective symbols Ga and Gu. Alternating and uniform graphs are partly characterized by the facts that, in them, each point and each edge forms part of at least one cycle of order n (hereafter called a *major cycle*), and that, in them, there occurs no cycle of order k where $2 < k < n$.

(a) Alternating partitions and their alternating graphs.

A Pa is defined as follows:

(1) A Pa contains no unit integers.

(2) Any integer in Pa occurs there an even number of times.

Hence every Pa is a Π_0.

A Ga is defined as follows:

(1) If $n = 2$, then $b(1, 2) = p_1$.

(2) If $n > 2$, then in Ga

$$b(i, i+1) = p_i - 1, \qquad i = 1, 3, 5, \cdots, (n-1),$$
$$b(j, j+1) = 1, \qquad j = 2, 4, 6, \cdots, (n-2), \qquad b(n, 1) = 1.$$

From these specifications, it follows that every heteropolar bond (if any) in Ga is single.

(b) Uniform partions and their uniform graphs.

A Pu is defined as follows:

(1) $p_1 = p_n$; (2) p_1 is even; n is odd.

A Gu is defined as follows:

(1) If $n = 1$, then Pu is not a Πu and Gu is the loop-rosette $(c_1, p_1/2, c_1)$.

(2) If $n > 1$, then $b(i, i+1) = p_1/2$, $i = 1, 2, \cdots, (n-1)$; $b(n, 1) = p_1/2$.

Here Pu is a Π_0 and Gu contains no heteropolar bond; moreover Gu contains no single bond unless $p_1 = 2$; in this latter case, all the bonds are single.

GENERAL RUBRIC FOR THE CLASSIFICATION OF ALL INITIAL Π'S.

	Class
$Fr(\Pi_0) = 0$	A
$Fr(\Pi_0) > 0$	
$p_1 = 2$	B
$p_1 > 2$	
$n = 3$	C
$n > 3$	
$p_1 = p_n$	D
$p_1 > p_2 = p_n$	E
$p_1 = p_{n-1} > p_n$	F
$p_1 = p_d > p_{d+1} = p_n;\ p_1 = p_2;\ p_{n-1} = p_n$	G
$p_1 > p_d > p_n;\ \Pi_0 = \Pi a + Pw$ [12]	
$\Sigma Pw = 0$	H
$\Sigma \Pi a = 0$	J
$\Sigma \Pi a > 0 < \Sigma Pw$	
$Fr(Pw) \overset{=}{>} 0$	K
$Fr(Pw) < 0$	L

Most of the classes of initial Π's given in the above rubric require still further division into subclasses before they can be successfully treated. In fact, the most feasible scheme yet found requires 30 such subclasses. Considerations of space make it out of the question to give the necessity proof for each of these. Nor is it essential to give the necessity proof for Theorem (2) in full. Six of the simpler proofs for individual subclasses are given hereafter. The methods used in the proofs for the remaining 24 subclasses are similar to those here shown (for subclasses A_3, C_1, D, E_1, G_1 and J_1) though in certain instances slightly more intricate.[13]

---

[12] The sum of m partitions is defined as follows: If all the integers of P are divided into m disjoint subsets which respectively constitute the partitions $P_1, \cdots, P_m$, then $P = P_1 + P_2 + \cdots + P_m$. A Pw is defined as a partition in which no two integers are alike.

It should be remembered that, in the six following necessity proofs for individual classes of Π's, the specifications given for each class apply *only* to the initial Π's of that class. The phrase "expand with respect to (i, j)" which is used in some of the constructions means that, in the indicated expansion, the cyclic edge to be expunged is one in the bond (i, j).

Subclass A_3. $Fr(\Pi_0) = 0, p_3 > 1$.

Π_0 is of class I, and G_0 is therefore the unique graph given (p. 678) for a Π of that class. To construct G_1, expand G_0 with respect to $(1, 2)$. To construct G_1', perform a neutral reversion on G_0 to form a graph in which $b(2, 3) = 1$ and $b(1, 1) = 1$; then expand this graph with respect to the unique loop-edge to form G_1'. G_1' contains an edge $(2, 3)$ terminated by two polyvalent points neither of which has the valence p_1. The tree $(c_2, 1, c_3)$ in G_1' is a $\bar{V}$. Thus the only unigraphic Π's in subclass A_3 are the initial ones.

Subclass C_1. In $\Pi_0, p_1 = p_3 > 2, n = 3$.

Π_0 is of class II, and G_0 is therefore the unique graph given (p. 678) for a Π of that class. Furthermore, every final Π in subclass C_1 is of class VII and hence has the unique graph given (p. —) for a Π of that class. In the unique G_0, $b(1, 2) = b(2, 3) = b(3, 1) = p_1/2$. To construct G_1, expand G_0 with respect to $(2, 3)$. To construct G_1', perform a neutral reversion on G_0 to form a graph in which $b(1, 1) = 1$; then expand this graph with respect to the unique loop-edge. G_1' is distinct from G_1 because, in G_1', the point c_1 is attached to two end-points, whereas in G_1 no point is attached to two end-points. Expand both G_1 and G_1' by $p_1/2 - 1$ successive expansions with respect to $(2, 3)$. Throughout this expansion, each G_i is distinct from the corresponding G_i' because G_i has a polyvalent point which is not attached to any end-point, whereas G_i' has no such polyvalent point. Any cycle of order 3 in $G'_{p_1/2}$ is a $\bar{V}$. Hence the only unigraphic Π's in subclass C_1 are the initial and final ones.

Class D. In $\Pi_0, p_1 = p_n > 2, n > 3$.

G_0 is either the alternating or the uniform graph for Π_0. To obtain G_0', construct for the partition $p_1, \cdots, p_4$ the component H_1 in which each point terminates two single bonds and one bond of breadth $p - 2$. If $n = 4$, H_1 is G_0'. If $n > 4$, the component H_2 is either the alternating or the uniform graph for the partition $p_5, \cdots, p_n$. If $n = 5$, transfuse each loop-edge in H_2 with an edge in a multiple bond in H_1 to form G_0'. If $n > 5$, transfuse one edge in H_2 with an edge in H_1 to form G_0'. In any case, any cycle of order 3 in G_0' is a $\bar{V}$. Hence there are no unigraphic Π's in class D.

---

[13] Any reader interested in the full necessary proof for all 30 subclasses may obtain a copy of that proof by application to the author.

Subclass E_1. $Fr(\Pi_0) = 1, n > 3, p_1 > p_2 = p_n$.

Π_0 is of class V, and G_0 is therefore the unique graph given (p. 679) for a Π of that class. G_0 is defined as follows: $b(1,2) = b(1,3) = p_2 - 1$, $b(2,3) = 1$, $b(1,i) = p_2$, $i = 4, \cdots, n$. To obtain G_1, expand G_0 with respect to $(1,2)$; to obtain G_1', expand G_0 with respect to $(2,3)$. In G_1, the tree $(c_2, 1, c_3)$ containing a homopolar edge is a $\bar{V}$. Hence the only unigraphic Π's in subclass E_1 are the initial ones.

Subclass G_1. In $\Pi_0, p_1 = p_d > p_{d+1} = p_n, p_1 = p_2, p_{n-1} = p_n = 2$.

The component I_1 is either the alternating or the uniform graph for the partition $p_1, \cdots, p_d$. The component I_2 is the loop-rosette $(c_n, 1, c_n)$. To construct the component H_1, transfuse the loop-edge in I_2 with an edge in $(1,2)$ in I_1. The component H_2 is either the alternating or the uniform graph for the partition $p_{d+1}, \cdots, p_{n-1}$. To construct G_0, transfuse an edge in H_2 with the single edge $(1,n)$ in H_1. To construct G_0', transfuse an edge in H_2 with an edge in $(1,2)$ in H_1. Any cycle of order n in G_0 contains only two heteropolar bonds and is therefore a $\bar{V}$. Hence there are no unigraphic Π's in subclass G_1.

Subclass J_1. $Fr(\Pi_0) > 0, p_g > p_{g+1}, g = 1, \cdots, (n-1), n > 3;$

$$p_1 = (p_2 + p_3 + \cdots + p_n) - 2z, 0 < 2z < p_n.$$

Attach each point c_g $(g = 2, \cdots, (n-1))$ to c_1 by p_g edges. Attach c_n to c_1 by $p_n - 2z$ edges and complete a (not loopless) graph for Π_0 by making c_n the loop-point for z dependent loop-edges. To construct G_0, transfuse each loop-edge dependent from c_n with an edge in $(1,3)$. To construct G_0', transfuse each loop-edge dependent from c_n with an edge in $(1,2)$. Any tree $(c_3, 1, c_n)$ in G_0 is a $\bar{V}$. Hence there are no ungraphic Π's in subclass J_1.

The full necessity proof (extracts from which have just been given) completes the proof for Theorem (2).

From Theorem (2) it follows that there are just four kinds of complete expansion sequence:

(1) Those in which no Π is unigraphic.

(2) Those in which just one, the initial Π is unigraphic.

(3) Those in which just two (the initial and the final) Π's are unigraphic.

(4) Those in which more than two (and here all) of the Π's are unigraphic.

LEMMA. *If any initial* Π *is multigraphic, then each of its expansion-products is multigraphic. If any intermediate* Π *is unigraphic, then each of its expansion and contraction products is unigraphic.*

Part IV. Polychromatic graphs.

Polychromatic graphs have been defined and the conditions for the identity of two such graphs have been given above. The existence condition (1. 4) for connected loopless graphs holds as well where the graph is polychromatic as where it is monochromatic. But for a graph with m colors ($m > 1$), this condition must be strengthened by one obvious further condition.

CONDITION (1. 5). $p_i = p_{i+m-1}$ for at least one p_i.

Where $m = 1$, condition (1. 5) reduces to the identity $p_i = p_i$. A Π which meets condition (1. 5) will be called a Π_m.

For every Π_m, there may be constructed at least one connected loopless graph involving each number of colors from 1 to m inclusive. And if, for Π_m ($m > 1$), there are x distinct connected loopless monochromatic graphs, the number of distinct connected loopless polychromatic graphs involving y colors ($2 \overline{\overline{<}} y \overline{\overline{<}} m$) must be $\overline{\overline{>}} x$. Hence, if Π has no unique monochromatic connected loopless graph, it can have no unique polychromatic connected loopless graph. To obtain uniqueness conditions for connected loopless polychromatic graphs it is therefore sufficient to determine how far the uniqueness conditions for connected loopless monochromatic graphs continue to hold in the polychromatic case. The various classes of Π's which are unigraphic in the monochromatic case (cf. above) are therefore next examined.

CLASS I. The Π's in this class have rosette graphs. No matter what the number of classes of like-valent blind points in a rosette, and no matter how the colors are distributed among these classes, such a graph must always be unique.

CLASS II. A graph for Π in this class must always be unique, no matter how the colors are distributed among the like-valent points, if any such are present. Trichromism is possible where $p_1 = p_2 = p_3$; dichromism is possible where $p_1 = p_2 > p_3$ or $p_1 > p_2 = p_3$.

CLASS III. A polychromatic graph for a Π in this class cannot be unique unless $p_1 = p_2$. Where this extra condition is met, two kinds of polychromism are possible: (a) The two polyvalent points are dichromatic. Here a polychromatic graph is unique if and only if all the univalent points are monochromatic. (b) The two polyvalent points are monochromatic. Here a polychromatic graph is unique if and only if one of the univalent points is of one color and all the other univalent points are of one different color.

CLASS IVa. A polychromatic graph for a Π of this class is unique if and only if one point is of one color and all the other points are of one different color.

CLASS IVb. In a graph for a Π of this class, polychromism among the bivalent points would destroy the uniqueness of the graph, but the two end points may be of different colors without destroying this uniqueness.

CLASS V. In a graph for a Π of this class, polychromism is possible only among the points of valence p_2. But such polychromism always destroys the uniqueness of the graph, because the colors may be distributed in various ways with respect to the unique homopolar edge.

CLASS VI. In a graph for a Π of this class, polychromism is possible only among the points of valence p_1. But such polychromism always destroys the uniqueness of the graph because the colors may be distributed in various ways with respect to the unique end-point.

CLASS VII. In a graph for a Π of this class, polychromism is possible both among the univalent and among the polyvalent points. But in either case, polychromism destroys the uniqueness of the graph.

The facts just stated may be summed up in the following lemmas.

LEMMA (3.1). *All polychromatic graphs of classes* I *and* II *are unique.*

LEMMA (3.2). *No polychromatic graph of any one of the classes* V, VI *or* VII *is unique.*

LEMMA (3.3). *A polychromatic graph of class* III *or class* IV *is unique only under the special conditions just described; such a unique graph can involve no more than two different colors.*

Lemmas (3.1)-(3.3) taken together imply theorem (3) which is thus proved.

GEORGE HERBERT JONES LABORATORY, UNIVERSITY OF CHICAGO.

ON SYSTEMS OF LINEAR DIFFERENTIAL EQUATIONS.[1]

By WILLIAM G. LEAVITT.

1. A normal form.

1. A normal form. A system of n linear homogeneous differential equations linear in a parameter λ may be written in vector form

$$U' = (\lambda A + P)U, \tag{1}$$

with U a column vector and A and P n-square matrices. The transformation $U = TU_1$ by a unimodular matrix T is easily seen to result in an equation in U_1 of form (1), in which the coefficient of λ is $T^{-1}AT$. It is known [1] that if the elements of A and its characteristic roots are holomorphic in a closed bounded region R, then there exists a matrix T unimodular in R such that $T^{-1}AT$ is in triangular form.

It may be remarked that the theorem is also true for the case R is unbounded.[2] It is easily verfied that the method of [1] may be extended to this case without significant alteration. The only possible difficulty is the requirement [1; p. 470] of a function whose finite expansion[3] is specified at a set of points which in this case may not be finite (although without a finite accumulation point). However, the existence of such a function is established by a theorem of Mittag-Leffler [2; pp. 5-6]. Another proof which may be extended to the unbounded case is that given in [3] for general principal ideal rings. It is known that the ring of all functions holomorphic in an unbounded region satisfies all the postulates of a principal ideal ring except the infinite chain condition [Cf. 4; p. 351]. However, the proof of [3] makes no use of the chain condition and so may be applied immediately.

It will accordingly be assumed that the coefficients of the original equation satisfy the above conditions. In the following it will be supposed that the transformation has already been performed, so that in (1) A is triangular.

2. A canonical form for second order equations. Consideration will be limited henceforth to the case $n = 2$. Since A is triangular, a further transformation $U = \exp[\frac{1}{2}\lambda \int (a_{11} + a_{22})\,dz]U_2$, with a_{11} and a_{22} the diagonal

---

[1] Received July 26, 1950. Presented (in part) to the American Mathematical Society, April 30, 1949.

[2] A function is said to be holomorphic in an unbounded region R if it is holomorphic in and on the boundary of every finite portion of R.

[3] The expansion to a finite number of terms.

elements of A, results in an equation in U_2 with λ-coefficient a matrix whose diagonal elements are the negative of each other. Again assuming the equation already transformed, A may be written

$$A = \begin{pmatrix} \mu & a_{12} \\ 0 & -\mu \end{pmatrix}.$$

Consider the equation $AT = TB$, where T is unimodular and

$$B = \begin{pmatrix} \mu & \psi \\ 0 & -\mu \end{pmatrix}. \tag{2}$$

Clearly T must be triangular (except in the trivial case $A = B = 0$) and, from the unimodularity condition, its diagonal elements t_{11} and t_{22} must be units. It then follows that

$$\psi = (a_{12}t_{22} + 2\mu t_{12})/t_{11}, \tag{3}$$

with t_{12} arbitrary. Thus ψ is any associate of any member of the modular class of $a_{12} \bmod \mu$.

Now it is known [4; p. 351] that there is an integral function γ such that $\gamma a_{12} \equiv d \bmod \mu$, where d is the G. C. D. of a_{12} and μ. It is easily seen that an exponential function can be constructed having, at a given point, any specified finite expansion with non-zero constant term, by properly choosing the corresponding finite expansion of its exponent. Since γ is prime to μ, it is accordingly evident (using the theorem of Mittag-Leffler cited above) that an exponential function exists whose expansion at each zero of μ agrees with that of γ to the order of the zero [See also 2; p. 56]. Thus a unit function exists which is congruent to $\gamma \bmod \mu$ and which is therefore usable in place of γ. It follows that the elements of T can always be chosen in (3) to make $\psi = d$. In case d is an associate of μ (that is, μ a divisor of a_{12}) ψ may from (3) be chosen equal to zero. Thus in all cases a form exists with ψ either zero or a proper divisor of μ. Conversely, if a_{12} is a proper divisor of μ, then from (3) if ψ is also a divisor of μ, it must be an associate of a_{12}. Also if $a_{12} = 0$, then ψ cannot be a proper divisor of μ, and so $\psi = 0$. Thus in either case ψ is unique except for associates. The transformed equation whose λ-coefficient is the matrix of (2), with ψ either zero or a proper divisor of μ, will be termed the *canonical form* of the equation.

3. Asymptotic solutions. The expressions to which solutions of equations of form (1) are asymptotic for large values of $|\lambda|$ are known only when the characteristic roots of A are unequal throughout R.[4] For second order equations, in the above notation, this would imply that μ is a unit function. It is to be expected that even for second order equations the problem of finding asymptotic solutions in general will be quite complicated. In this section is to be considered the simplest case, in which it is assumed that R is bounded and that μ has a single zero of the first order (which can be taken as the origin without loss of generality). It will be assumed that the equation has been reduced to canonical form, and so, according to the results of the last section, ψ has one of the two forms 0 or 1.

---

[4] See [5]. In applying the method to the complex case certain additional assumptions on R may be needed to ensure convergence.

Case I. $\psi = 1$. The equation to be solved has form (1) with

$$A = \begin{pmatrix} \mu & 1 \\ 0 & -\mu \end{pmatrix}.$$

Since it is assumed that R is bounded, the further transformation,

$$U = (\lambda + p_{12})^{\frac{1}{2}} \exp\left[\tfrac{1}{2}\int_0^z (p_{11} + p_{22})\,dz\right] Y,$$

is non-singular for $|\lambda|$ sufficiently large. Equation (1) then becomes

$$Y' = (\lambda A + Q) Y \tag{4}$$

where

$$Q = \begin{pmatrix} p - r & p_{12} \\ p_{21} & -p - r \end{pmatrix}, \text{ with } \begin{cases} p = \tfrac{1}{2}(p_{11} - p_{22}), \\ r = p'_{12}/2(\lambda + p_{12}). \end{cases}$$

The elimination of y_2 from the component equations of (4) is found to lead to an equation of form

$$y_1'' = [\lambda^2\mu^2 + \lambda\chi_1(z) + \chi_2(z, \lambda)] y_1, \tag{5}$$

where χ_1 and χ_2 are holomorphic in R and χ_2 is bounded for $|\lambda|$ sufficiently large.

This is an equation of the type treated by Langer [6]. Thus with the added postulates [6; (i) p. 93 (second part) and (iv) p. 100] the expressions to which the solutions of (5) are asymptotic for large $|\lambda|$ are obtainable. The first component equation of (4) then furnishes a solution for y_2 and, by reversing the transformation leading to (4), asymptotic solutions are obtained for the original equation.

Case II. $\psi = 0$. Transform (1) by

$$U = \begin{bmatrix} \exp\int_0^z (\lambda\mu + p_{11})\,dz & 0 \\ 0 & \exp\int_0^z (-\lambda\mu + p_{22})\,dz \end{bmatrix} Y.$$

Equation (1) then becomes

$$(6)\qquad Y' = \begin{bmatrix} 0 & p_1 \exp(-\xi) \\ & \\ p_2 \exp(\xi) & 0 \end{bmatrix} Y, \text{ where } \begin{cases} \xi = 2\lambda \int_0^z \mu dz \\ p_1 = p_{12} \exp \int_0^z (p_{22} - p_{11}) dz \\ p_2 = p_{21} \exp \int_0^z (p_{11} - p_{22}) dz. \end{cases}$$

For some constant α, define recursively

$$(7)\qquad \begin{aligned} &y_{11}^{(i)} = \int_\alpha^z p_1 \exp(-\xi) y_{21}^{(i)} dz && (i = 0, 1, \cdots) \\ &y_{21}^{(0)} = 1, \quad y_{21}^{(i)} = \int_\alpha^z p_2 \exp(\xi) y_{11}^{(i-1)} dz && (i = 1, 2, \cdots), \end{aligned}$$

then the expressions

$$(8)\qquad y_{11} = \sum_0^\infty y_{11}^{(i)}, \qquad y_{21} = \sum_0^\infty y_{21}^{(i)},$$

are the components of a formal solution of (6).

Certain conditions are now to be assumed to ensure convergence:

(i) For some K and $\epsilon > 0$ there exists a curve joining the origin with every z for which $|z| < \epsilon$, such that $\int_0^z |dz| < K|z|$. Call the set of all such curves the class c.

(ii) The function $\int_0^z \mu dz$ is zero in R only at the origin.

In the region R_ξ corresponding to R define (in general four) sectors S_j by $-\pi/2 + j\pi \leqq \arg \xi \leqq \pi/2 + j\pi$, where $j = 0, 1, 2, 3$. Then assume, finally,

(iii) Each sector contains a point α_j which may, for each z of S_j, be connected with the origin by an ordinary curve C_z through z on which the real part $\Re(\xi)$ is monotonic. Call the set of all such curves the class C, and let $L(z)$ be the length of C_z. It will be assumed, further, that $L(z)$ is a bounded function of z over R.

Let H be an upper bound for the absolute values of the functions p_1, p_2, μ, their derivatives, $L(z)$, and $2\int_0^z \mu dz/\mu^2$. To specify the constant α in

(7) let α be α_j or 0 according as $j = 0, 2$ or $j = 1, 3$. Then for any z_1 on the curve between α and z, $\Re(\xi_1) \geqq \Re(\xi)$.

Suppose first that $|\xi| \leqq N$. By (ii) and the fact that μ has a zero of the first order at the origin, $|\int_0^z \mu dz| > |z|^2 K_1$ for some K_1. Thus $|z| < (N/K_1|\lambda|)^{\frac{1}{2}}$, and so $|\lambda| > N/\epsilon^2 K_1$ implies $|z| < \epsilon$. In sectors S_1 and S_3 $\alpha = 0$ in (7) and so the integration may be taken over a curve of class c. It is then easy to show by induction that $|y_{11}^{(i)}| \leqq M^{2i+1}|\lambda|^{-i-\frac{1}{2}}$, $|y_{21}^{(i)}| \leqq M^{2i}|\lambda|^{-i}$, where $M = HK(N/K_1)^{\frac{1}{2}} \exp N$. Hence the series in (8) converge when λ also satisfies the condition $|\lambda| > M^2$.

If $|\xi| > N$, again for sectors S_1 and S_3, the integration is taken over a curve C_z of class C. Since this is an ordinary curve, there is a last point β on the curve for which $|\xi| = N$, and the part of C_z from 0 to β may be replaced by a curve of class c. Then from (7), integrating by parts,

$$(9)\quad y_{11}^{(0)} = \int_0^\beta p_1 \exp(-\xi)\, dz + p_1 \exp(-\xi)/(-2\lambda\mu)\Big|_\beta^z$$
$$+ 1/(2\lambda) \int_\beta^z (p_1/\mu)' \exp(-\xi)\, dz.$$

The first integral is of the type considered above for $|\xi| \leqq N$, and so is less in absolute value than $M/|\lambda|^{\frac{1}{2}}$. In the above, $|\lambda|$ was chosen to make this ratio less than 1. It is now to be required that $|\lambda|$ also be chosen large enough to make $M/|\lambda|^{\frac{1}{2}} < H^4/N$. Now we may write $\lambda\mu^2 = \xi\mu^2/2 \int_0^z \mu dz$, and since on the curve from β to z $|\xi| > N$, it follows that $1/|\lambda\mu^2| < H/N$. Also by (iii), for any z_1 on this curve, $\Re(\xi_1) \geqq \Re(\xi)$, so that $|\exp(-\xi)| \geqq |\exp(-\xi_1)|$. It may then be verified (assuming $H > 1$) that $|y_{11}^{(0)}| < 5H^4 |\exp(-\xi)|/N$. Thus $y_{11}^{(0)} = E \exp(-\xi)/N$ where E is a generic symbol for any function for which $|E| < 5H^4$.

The curve of integration for $y_{21}^{(1)}$ in (7), for sectors S_1 and S_3, may be split in similar fashion. On the curve of class c from 0 to β it was shown above that $|y_{11}^{(0)}| < M/|\lambda|^{\frac{1}{2}}$. It is then seen that $y_{21}^{(1)}| < H^8/N^2 + 5H^6/N$, so that $y_{21}^{(1)} = E^2/N$.

This process may be continued, integrating by parts for each $y_{11}^{(i)}$. Then, using the fact that $y_{21}^{(i)\prime} = p_2 \exp(\xi) y_{11}^{(i-1)}$, it may be shown by induction that $y_{11}^{(i)} = E^{2i+1} \exp(-\xi)/N^{i+1}$ and then $y_{21}^{(i)} = E^{2i}/N^i$. Accordingly with a choice first of $N > (5H^4)^2$, and then $|\lambda|$ as specified above, the series in (8) converge and represent the solution.

In the sectors S_0 and S_2 the situation is somewhat less simple. The choice must now be $\alpha = \alpha_j$, so that again $|\exp(-\xi)| \geqq |\exp(-\xi_1)|$ for any z_1 on the curve C_z of class C joining α_j with z. If there are points on C_z for which $|\xi_1| \leqq N$, the portion of the curve joining the first and last points on the curve for which $|\xi_1| = N$ may be replaced by two curves (by way of the origin) of class c. The remainder of C_z consists of one or two segments of a curve of class C on which $|\xi_1| > N$. Thus clearly an analysis similar to that above is possible showing that for $|\lambda|$ sufficiently large the series in (8) converge. Accordingly there exist solutions, in any sector, $y_{11} \sim y_{11}^{(0)}$ and $y_{21} \sim 1$.[5]

For a second solution take

$$y_{22}^{(i)} = \int_{\alpha}^{z} p_2 \exp(\xi) y_{12}^{(i)} dz, \qquad (i = 0, 1, \cdots)$$

$$y_{12}^{(0)} = 1, \quad y_{12}^{(i)} = \int_{\alpha}^{z} p_1 \exp(-\xi) y_{22}^{(i-1)} dz \qquad (i = 1, 2, \cdots),$$

with $\alpha = \alpha_j$ if $j = 1, 3$ and $\alpha = 0$ if $j = 0, 2$. If the curve of integration is of class C, then $\Re(\xi_1) \leqq \Re(\xi)$ for any z_1 on the curve, so that the analysis evidently parallels that above. There thus exists a second solution $y_{12} \sim 1$ and $y_{22} \sim y_{22}^{(0)}$.

Let Y_j designate a matrix of solutions (each column a solution) having the above form in sector S_j, then

$$Y_0 \sim \begin{bmatrix} \int_{\alpha_0}^{z} p_1 \exp(-\xi)\, dz & 1 \\ 1 & \int_{0}^{z} p_2 \exp(\xi)\, dz \end{bmatrix}.$$

From (6), $Y_0 K$ is a matrix of solutions for any constant matrix K, and so it represents the general solution. Thus each $Y_i = Y_0 K_i$ for some constant matrix K_i. Now the origin is a point common to all sectors, and hence it may be used to evaluate the constants K_i. Thus since $|Y_0| \sim -1$, it follows that $Y_i = Y_0 Y_0^{-1}(0) Y_i(0)$. The solution for the original equation may then be obtained by reversing the transformation leading to (6).

UNIVERSITY OF NEBRASKA.

---

[5] The symbol $\sim$ indicates expressions asymptotic for large $|\lambda|$. It is easy to show, also, that if $|\exp(-\xi)|$ is sufficiently large (that is, if the point is sufficiently removed from the curve $\Re(\xi) = 0$), then for $|\xi| > N$, $y_{11}^{(0)}$ is approximated by $-p_1 \exp(-\xi)/2\lambda\mu$.

BIBLIOGRAPHY.

[1] W. G. Leavitt, "A normal form for matrices whose elements are holomorphic functions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 463-472.

[2] G. Mittag-Leffler, "Sur la représentation analytique des fonctions monogènes uniformes," *Acta Mathematica*, vol. 4 (1884), pp. 1-79.

[3] W. G. Leavitt and G. Whaples, "On matrices with elements in a principal ideal ring," *Bulletin of the American Mathematical Society*, vol. 55 (1949, pp. 117-118.

[4] O. Helmer, "Divisibility properties of integral functions," *Duke Mathematical Journal*, vol. 6 (1940), pp. 345-356.

[5] G. D. Birkhoff and R. E. Langer, "The boundary problems and developments associated with a system of ordinary linear differential equations of the first order," *Proceedings of the American Academy of Arts and Sciences*, vol. 58 (1923), pp. 52-128.

[6] R. E. Langer, "The asymptotic solutions of certain linear ordinary differential equations of the second order," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 90-106.

ISOMORPHIC GROUPS OF LINEAR TRANSFORMATIONS, II.* [1]

By C. E. Rickart.

Introduction. In a previous paper [7] we have discussed the structure of isomorphisms of certain groups of linear transformations which reduce in the finite-dimensional case to full linear groups. In the present paper a similar investigation is carried out for groups of linear transformations which reduce in finite dimensions to unitary (including orthogonal) or symplectic groups. As in the previous case, it turns out that the group isomorphisms are essentially generated by isomorphisms of the underlying vector spaces on which the linear transformations operate.

In the finite-dimensional case, similar results for automorphisms have been announced by Dieudonné [2, 3].[2] Dieudonné assumes throughout that the coefficient domain is a field, that the index of the vector space is not zero, and, in the orthogonal and unitary cases, that the field is not of characteristic two. We assume the characteristic to be different from two throughout the discussion. However, in the unitary case, the coefficient domain can be any division ring with an involution, except that the characteristic must be different from two and the case of a field F_3 with exactly three elements is excluded. There is also no restriction here on the index of the vector spaces.

In view of Dieudonné's results, the main force of our discussion is obtained in the infinite-dimensional case. Therefore we have assumed throughout that the vector spaces have dimension at least equal to six. This avoids the special situations which occur for low dimensions and makes it possible to give a relatively uniform treatment of the whole problem. For similar reasons, we have not attempted a discussion of the symplectic case for characteristic two nor the unitary case for the field F_3.

In §1 vector spaces with a scalar product (called self-dual spaces) and self-adjoint involutions on such spaces are discussed. These spaces include

---

* Received July 24, 1950.

[1] Presented to the Society, April, 1950. We take this opportunity to express our appreciation to N. Jacobson for his encouragement and advice during the writing of this paper.

[2] Details of the proofs of these announced results will presumably appear in [1]. L. K. Hua [5] has also obtained the result for the finite-dimensional symplectic case when the field is not of characteristic two. (Added in proof.) Reference [1] appeared after this paper was written. The methods used there are, in general, quite different from ours.

the symplectic and unitary spaces as special cases. In §§2 and 3 the special problems which arise in the unitary and symplectic cases respectively are discussed. The representation theorem for the group isomorphism is proved in §4.

1. Self-dual linear vector spaces. Let $\mathfrak{X}$ be a left linear vector space over a division ring $\mathfrak{D}$ and let $\lambda \rightarrow \lambda^*$ be a given involution in $\mathfrak{D}$. In other words, $(\lambda^*)^* = \lambda$, $(\lambda + \mu)^* = \lambda^* + \mu^*$, and $(\lambda\mu)^* = \mu^*\lambda^*$. The space $\mathfrak{X}$ is said to be *self-dual* provided there exists a *scalar product* (x, y) defined on $\mathfrak{X} \times \mathfrak{X}$ to $\mathfrak{D}$ possessing the following properties:

(i) $(\lambda x + \mu y, z) = \lambda(x, z) + \mu(y, z)$,

(ii) $(x, y) = 0$ *for all* x *implies* $y = 0$,

(iii) $(y, x) = \epsilon(x, y)^*$, *where* $\epsilon = \pm 1$ *and is independent of* x, y.

If $\mathfrak{X}^*$ is the right linear vector space over $\mathfrak{D}$ whose elements are identical with those of $\mathfrak{X}$, with addition defined as in $\mathfrak{X}$ and multiplication by scalars defined by $x\lambda = \lambda^* x$, then $\mathfrak{X}$ and $\mathfrak{X}^*$ are dual linear vector spaces relative to (x, y) as defined by Jacobson [6, p. 15]. An important consequence is that, if $x_1, \cdots, x_n$ are linearly independent elements of $\mathfrak{X}$, then there exist elements $y_1, \cdots, y_n$ in $\mathfrak{X}$ such that $(x_i, y_j) = \delta_{ij}$, where $\delta_{ij} = 0$ for $i \neq j$ and $\delta_{ii} = 1$ $(i, j = 1, 2, \cdots, n)$.

For reasons which will appear later, it will be assumed throughout our discussion that $\mathfrak{D}$ is not of characteristic two and furthermore, in case $\mathfrak{X}$ is unitary (as defined below), that $\mathfrak{D}$ is not equal to the field F_3 which contains exactly three elements.

Two vectors x, y in $\mathfrak{X}$ are said to be *orthogonal* provided $(x, y) = 0$. Two subspaces $\mathfrak{M}$, $\mathfrak{N}$ are said to be orthogonal provided $(x, y) = 0$ for all $x \in \mathfrak{M}$ and $y \in \mathfrak{N}$. In this case we write $\mathfrak{M} \perp \mathfrak{N}$. If $\mathfrak{M}$ is an arbitrary subspace of $\mathfrak{X}$, then $\mathfrak{M}^{\perp}$ will denote the set of all $x \in \mathfrak{X}$ such that $(x, y) = 0$ for every $y \in \mathfrak{M}$. We call $\mathfrak{M}^{\perp}$, which is a linear subspace of $\mathfrak{X}$, the *orthogonal complement*[3] of $\mathfrak{M}$. A subspace $\mathfrak{M}$ in $\mathfrak{X}$ is said to be *isotropic* if $\mathfrak{M} \cap \mathfrak{M}^{\perp} \neq (0)$ and is said to be *totally isotropic* if $\mathfrak{M} \subseteq \mathfrak{M}^{\perp}$. The maximum dimension which a totally isotropic subspace of $\mathfrak{X}$ can have is called the *index* of $\mathfrak{X}$ [4, p. 17]. A subspace of $\mathfrak{X}$ which is non-isotropic is itself a self-dual space relative to the scalar product (x, y) restricted to the subspace. A non-zero vector $u \in \mathfrak{X}$ such that $(u, u) = 0$ is called an *isotropic*

---

[3] $\mathfrak{M}^{\perp}$ is not in general an algebraic complement of $\mathfrak{M}$ in the usual sense. In fact, even if $\mathfrak{M} \cap \mathfrak{M}^{\perp} = (0)$, we need not have $\mathfrak{X} = \mathfrak{M} \oplus \mathfrak{M}^{\perp}$ when $\mathfrak{X}$ is infinite-dimensional However, see Lemma 1.1 below.

vector. If u is an arbitrary isotropic vector in $\mathfrak{X}$, then there exists an isotropic vector $v \varepsilon \mathfrak{X}$ such that $(u, v) = 1$. In fact, let w be any vector in $\mathfrak{X}$ such that $(u, w) = 1$ and define $v = w - \frac{1}{2}(w, w)u$. Then $(u, v) = 1$ and $(v, v) = 0$. The vectors u, v are linearly independent and the two-dimensional subspace of $\mathfrak{X}$ spanned by u, v is non-isotropic. Thus every isotropic vector can be embedded in a non-isotropic plane.

A self-dual space in which every vector is isotropic is said to be *symplectic.* In this case it is not difficult to show that $(y, x) = -(x, y)$, for all x, y, and $\lambda^* = \lambda$, for all $\lambda \varepsilon \mathfrak{D}$, so that $\mathfrak{D}$ must be a field. It is well-known that the dimension of a symplectic space, if finite, is even.

A self-dual space which contains non-isotropic vectors is said to be *unitary.* If $\epsilon = -1$ in this case, then, by a trivial modification of the involution and scalar product, it is possible to obtain[4] $\epsilon = 1$. Therefore we shall always assume $\epsilon = 1$ in the unitary case. A non-isotropic subspace of a unitary space obviously cannot be symplectic and so accordingly must itself be unitary. Let $\mathfrak{M}$ be a finite-dimensional non-isotropic subspace of a unitary space $\mathfrak{X}$. Then there exists a basis $u_1, \cdots, u_n$ for $\mathfrak{M}$ consisting of non-isotropic orthogonal vectors; that is, $(u_i, u_j) = 0$ for $i \neq j$ and $(u_i, u_i) \neq 0$. This is proved as follows. Choose u_1 as any non-isotropic vector in $\mathfrak{M}$ and suppose $u_1, \cdots, u_k$, where $1 \leq k < n = \dim \mathfrak{M}$, already chosen in $\mathfrak{M}$ such that $(u_i, u_j) = 0$ for $i \neq j$ and $(u_i, u_i) \neq 0$. Observe that $u_1, \cdots, u_k$ must be linearly independent and span a non-isotropic subspace of $\mathfrak{M}$. Set[5] $\mathfrak{N} = \mathfrak{D}u_1 + \cdots + \mathfrak{D}u_k$ and note that $\dim (\mathfrak{M} \cap \mathfrak{N}^{\perp}) = n - k$. Since $k < n$ and[5a] $\mathfrak{M} = \mathfrak{N} \oplus (\mathfrak{M} \cap \mathfrak{N}^{\perp})$, $\mathfrak{M} \cap \mathfrak{N}^{\perp}$ is non-zero and non-isotropic. Choose a non-isotropic vector u_{k+1} in $\mathfrak{M} \cap \mathfrak{N}^{\perp}$. Then $(u_i, u_{k+1}) = 0$ for $i = 1, \cdots, k$ and $(u_{k+1}, u_{k+1}) \neq 0$. The desired result now follows by induction. If $\mathfrak{X}$ is unitary and $\mathfrak{D}$ is a field with $\lambda^* = \lambda$ for all λ, then $\mathfrak{X}$ is said to be an *orthogonal* space.

Now return to a general self-dual space $\mathfrak{X}$ (that is, $\mathfrak{X}$ is either symplectic or unitary) and let A be a linear transformation on $\mathfrak{X}$. A second linear transformation A^* on $\mathfrak{X}$ such that $(xA, y) = (x, yA^*)$ for all x, y is called the *adjoint* of A. If A^* exists, then it is unique. In case $A^* = A$, then A is said to be *self-adjoint.* A linear transformation P such that $P^2 = P$ is called a *projection.* If $P^* = P$, then the range of P is a non-isotropic subspace of $\mathfrak{X}$. In fact, suppose $y \varepsilon (\mathfrak{X}P) \cap (\mathfrak{X}P)^{\perp}$. Then, for all x, we have $(x, y) = (x, yP) = (xP, y) = 0$. Therefore $y = 0$ and $\mathfrak{X}P$ is non-isotropic. Similarly $\mathfrak{X}(I - P)$ is non-isotropic, where I is the identity transformation. Furthermore $\mathfrak{X}(I - P) = (\mathfrak{X}P)^{\perp}$ and therefore $\mathfrak{X} = (\mathfrak{X}P) \oplus (\mathfrak{X}P)^{\perp}$. Conversely, if $\mathfrak{X} = \mathfrak{M} \oplus \mathfrak{M}^{\perp}$ and P is the projection of $\mathfrak{X}$ onto $\mathfrak{M}$, then P is self-adjoint. Decompositions of the form $\mathfrak{X} = \mathfrak{M} \oplus \mathfrak{M}^{\perp}$ do not hold in general for non-isotropic subspaces $\mathfrak{M}$ which are infinite-dimensional. However, if $\mathfrak{M}$ is finite-dimensional the decomposition does hold.

---

[4] This fact was pointed out to us in conversation by N. Jacobson.

[5] The one-dimensional subspace of $\mathfrak{X}$ which contains a non-zero vector $u \varepsilon \mathfrak{X}$ will be denoted by $\mathfrak{D}u$.

[5a] See Lemma 1.1 below.

LEMMA 1.1. *If $\mathfrak{M}$ is a finite-dimensional non-isotropic subspace of $\mathfrak{X}$, then $\mathfrak{X} = \mathfrak{M} \oplus \mathfrak{M}^{\perp}$.*

Proof. Since $\mathfrak{M}$ is non-isotropic, it is self-dual relative to (x, y) restricted to $\mathfrak{M}$. Therefore, if $u_1, \cdots, u_n$ is a basis for $\mathfrak{M}$, there exist $v_1, \cdots, v_n \in \mathfrak{M}$ such that $(u_i, v_j) = \delta_{ij}$. Observe that $v_1, \cdots, v_n$ must be linearly independent and hence constitute a basis for $\mathfrak{M}$. Now, for $x \in \mathfrak{X}$, write $x = u + v$, where $u = \sum_{k=1}^{n} (x, v_k) u_k$, $v = x - u$. Then for each v_j, $(v, v_j) = (x, v_j) - \sum_{k=1}^{n} (x, v_k)(u_k, v_j) = 0$. Therefore $v \in \mathfrak{M}^{\perp}$. Since $u \in \mathfrak{M}$, we have $\mathfrak{X} = \mathfrak{M} \oplus \mathfrak{M}^{\perp}$.

In addition to the usual operations of addition and multiplication of linear transformations, we shall find it convenient to use the *circle operation*, $A \circ B = A + B - AB$. This operation is associative and has the identically zero transformation as an identity. Since $I - (A \circ B) = (I - A)(I - B)$, the mapping $A \to I - A$ takes the circle operation into ordinary multiplication. If a transformation A has an inverse relative to the circle operation, it will be denoted by A°. Thus $A^{\circ} \circ A = A \circ A^{\circ} = 0$ and $I - A^{\circ} = (I - A)^{-1}$, where $(I - A)^{-1}$ is the ordinary multiplicative inverse of $I - A$. We shall denote by $\mathfrak{U}(\mathfrak{X})$ the class of all linear transformations A on $\mathfrak{X}$ such that $A^* = A^{\circ}$. Observe that, if A° exists, then $A \in \mathfrak{U}(\mathfrak{X})$ if, and only if,

$$(xA, yA) = (xA, y) + (x, yA) \tag{1.1}$$

for all x, y. $\mathfrak{U}(\mathfrak{X})$ is a group under the circle operation and is isomorphic, under the correspondence $A \to I - A$, to the multiplicative group of all linear transformations T such that $T^* = T^{-1}$; that is $(xT, yT) = (x, y)$ for all x, y. The elements of $\mathfrak{U}(\mathfrak{X})$ are called *symplectic, unitary* or *orthogonal* transformations according as $\mathfrak{X}$ is symplectic, unitary or orthogonal respectively.

If F is a finite-dimensional linear transformation on $\mathfrak{X}$ for which F^* exists, then it is not difficult to verify that F and F^* have the forms [6]

---

[6] For a proof that F has the indicated form, see [6, p. 17].

$$xF = \sum_{i=1}^{n} (x, b_i) a_i, \qquad xF^* = \epsilon \sum_{i=1}^{n} (x, a_i) b_i,$$

where n is the dimension of $\mathfrak{X}F$.

A linear transformation T is called an *involution* provided $T^\circ = T$; that is, $T \circ T = 0$ or $T^2 = 2T$. An involution T is in $\mathcal{U}(\mathfrak{X})$ if, and only if, $T^* = T$. In this case $\frac{1}{2}T$ is a self-adjoint projection, so that $\mathfrak{M} = \mathfrak{X}T$ is a non-isotropic subspace and $\mathfrak{X} = \mathfrak{M} \oplus \mathfrak{M}^\perp$. $\mathfrak{M}$ is called the *subspace of* T. Observe that T is uniquely determined by $\mathfrak{M}$ and that $xT = 2x$ for every $x \varepsilon \mathfrak{M}$ while $xT = 0$ for $x \varepsilon \mathfrak{M}^\perp$. If $\mathfrak{M}$ has minimum dimension (that is, $\dim \mathfrak{M} = 2$ or 1 according as $\mathfrak{X}$ is symplectic or unitary), then T is said to be *minimal.* If $\mathfrak{M}^\perp$ has minimal dimension, then T is said to be *maximal.* If T is either minimal or maximal, then it is said to be *extremal.* If $\mathfrak{M}$ is the subspace of an involution T in $\mathcal{U}(\mathfrak{X})$, then $\mathfrak{M}^\perp$ is the subspace of $(2I) \circ T$. Hence T is minimal if, and only if, $(2I) \circ T$ is maximal. A one-dimensional involution U is in $\mathcal{U}(\mathfrak{X})$ if, and only if, it has the form $xU = 2(x, u)(u, u)^{-1}u$. It follows from Lemma 1.1 that, for any finite-dimensional non-isotropic subspace $\mathfrak{M}$ of $\mathfrak{X}$, there exists an involution in $\mathcal{U}(\mathfrak{X})$ with $\mathfrak{M}$ as its subspace. In fact, if P is the projection of $\mathfrak{X}$ onto $\mathfrak{M}$ given by Lemma 1.1, then $2P$ is the desired involution.

LEMMA 1.2. *Let T_1, T_2 be involutions in $\mathcal{U}(\mathfrak{X})$ with subspaces $\mathfrak{M}_1$, $\mathfrak{M}_2$ respectively. A sufficient condition for $T_1 \circ T_2 = T_2 \circ T_1$ is that $\mathfrak{M}_1$ be contained in either $\mathfrak{M}_2$ or $\mathfrak{M}_2^\perp$. If T_1 is minimal, the condition is also necessary.*

Proof. The proof of the sufficiency is not difficult so will be omitted.[7] Therefore assume T_1 to be minimal and that T_1, T_2 commute. (Note that $T_1 \circ T_2 = T_2 \circ T_1$ is equivalent to $T_1T_2 = T_2T_1$.) Since T_1, T_2 commute, $\frac{1}{4}T_1T_2$ is a self-adjoint projection with range $\mathfrak{M}_1 \cap \mathfrak{M}_2$. Therefore $\mathfrak{M}_1 \cap \mathfrak{M}_2$ is non-isotropic. Since $\mathfrak{M}_1$ is minimal, it follows that either $\mathfrak{M}_1 \cap \mathfrak{M}_2 = \mathfrak{M}_1$ or $\mathfrak{M}_1 \cap \mathfrak{M}_2 = (0)$. A similar argument applied to the projection $\frac{1}{2}T_1(I - \frac{1}{2}T_2)$ shows that either $\mathfrak{M}_1 \cap \mathfrak{M}_2^\perp = (0)$ or $\mathfrak{M}_1 \cap \mathfrak{M}_2^\perp = \mathfrak{M}_1$. This completes the proof.

COROLLARY. *Let T_1, T_2 be extremal involutions with associated minimal subspaces $\mathfrak{M}_1$, $\mathfrak{M}_2$. If T_1, T_2 commute, then either $\mathfrak{M}_1 = \mathfrak{M}_2$ or $\mathfrak{M}_1 \perp \mathfrak{M}_2$.*

LEMMA 1.3. *Let P be a self-adjoint projection in $\mathfrak{X}$ and let Z be an involution in $\mathcal{U}(\mathfrak{X})$ which commutes with every minimal involution $U \varepsilon \mathcal{U}(\mathfrak{X})$ such that $UP = PU = U$. Then $PZ = \lambda P$, where $\lambda = 0$ or 2.*

---

[7] See [7, Lemma 2.2] for a proof.

Proof. If $\mathfrak{X}$ is symplectic, then every non-zero vector $u \in \mathfrak{X}P$ can be embedded in a minimal non-isotropic subspace $\mathfrak{M} \subseteq \mathfrak{X}P$. By Lemma 1. 1, there exists a minimal involution $U \in \mathcal{U}(\mathfrak{X})$ with $\mathfrak{M}$ as its subspace. Obviously $PU = UP = U$ so that Z commutes with U. Now, by Lemma 1. 2, $\mathfrak{M}$ must be contained in either the subspace of Z or its orthogonal complement. It follows that $uZ = \lambda_u u$, where $\lambda_u = 0$ or 2. If u and v are any non-zero vectors in $\mathfrak{X}P$, the additivity of Z implies $(\lambda_{u+v} - \lambda_u)u = (\lambda_{u+v} - \lambda_v)v = 0$. Since the λ's assume only the values 0 and 2, it follows that $\lambda_u = \lambda_{u+v} = \lambda_v$. Therefort λ_u is a constant λ independent of u. This completes the proof for the symplectic case.

If $\mathfrak{X}$ is unitary, then the above proof gives $uZ = \lambda_u u$, where $\lambda_u = 0$ or 2, for every non-isotropic $u \in \mathfrak{X}P$. Hence let u be an isotropic vector in $\mathfrak{X}P$ and choose a second isotropic vector in $\mathfrak{X}P$ such that $(u, v) = 1$. Then $u + v$, $u - v$ are non-isotropic so that $(u + v)Z = \lambda_{u+v}(u + v)$ and $(u - v)Z = \lambda_{u-v}(u - v)$. We prove now that $\lambda_{u+v} = \lambda_{u-v}$. At this point it is necessary to use the hypothesis $\mathcal{D} \neq F_3$, the problem being to obtain an $\alpha \in \mathcal{D}$ such that $\alpha \neq \pm 1$ and $\alpha + \alpha^* \neq 0$. Since $\mathcal{D} \neq F_3$, there exists a non-zero $\beta \in \mathcal{D}$ such that $\beta \neq \pm 1$. If $\beta + \beta^* \neq 0$, take $\alpha = \beta$. If $\beta + \beta^* = 0$, take $\alpha = \beta + 1$. Then $\alpha + \alpha^* = 2$ and clearly $\alpha \neq \pm 1$. Thus the desired α exists. With this choice of α, the element $u + \alpha v$ is non-isotropic, so that $(u + \alpha v)Z = \lambda_{u+\alpha v}(u + \alpha v)$. On the other hand, if $\lambda_{u+v} \neq \lambda_{u-v}$, then either $uZ = u - v$ and $vZ = v - u$ or $uZ = u + v$ and $vZ = u + v$, according as $\lambda_{u+v} = 0$ or 2. Therefore $(u + \alpha v)Z = (1 - \alpha)(u - v)$ or $(u + \alpha v)Z = (1 + \alpha)(u + v)$ according as $\lambda_{u+v} = 0$ or 2. Since u, v are linearly independent, it follows that $\lambda_{u+\alpha v} = 1 - \alpha$ or $\lambda_{u+\alpha v} = 1 + \alpha$. But $\lambda_{u+\alpha v} = 0$ or 2 so that either possibility contradicts $\alpha \neq \pm 1$. It follows that $\lambda_{u+v} = \lambda_{u-v}$ and hence $(u + v)Z = \lambda_{u+v}(u + v)$ and also $(u - v)Z = \lambda_{u+v}(u - v)$. Adding these equations and dividing by two, we obtain $uZ = \lambda_{u+v}u$ or $uZ = \lambda_u u$, where $\lambda_u = 0$ or 2. Again it follows that λ_u is a constant λ independent of u and this completes the proof.

Lemma 1. 3 is not true in general without the condition[8] $\mathcal{D} \neq F_3$ and is the only place in our discussion where this condition is needed. On the other hand, an examination of the proof shows that $\mathcal{D} = F_3$ can be admitted if there are no isotropic vectors in $\mathfrak{X}P$ and in particular if $\mathfrak{X}$ has index zero. It is also not difficult to prove the lemma when $\mathcal{D} = F_3$ provided $\dim(\mathfrak{X}P) > 2$.

Consider now a subgroup $\mathcal{G}$ of the group $\mathcal{U}(\mathfrak{X})$ which contains all of the minimal involutions in $\mathcal{U}(\mathfrak{X})$. The objective in the remainder of this section is to obtain a characterization of the extremal involutions in $\mathfrak{G}$ in terms only of the group operation in $\mathfrak{G}$.

---

[8] See, e. g., [4, p. 24].

Let $\mathfrak{S}$ denote a set of involutions in $\mathfrak{G}$ and denote by $c(\mathfrak{S})$ the set of all involutions in $\mathfrak{G}$ each of which commutes with every element of $\mathfrak{S}$. Also, if T_1, T_2 are involution in $\mathfrak{G}$, denote by $\rho(T_1, T_2)$ the number of distinct involutions in $c(c(T_1, T_2))$. In addition, define

$$\rho = \max \rho(T_1, T_2), \qquad T_1 \circ T_2 = T_2 \circ T_1,$$

$$\rho_T = \max \rho(T, T'), \qquad T \circ 'T = T' \circ T.$$

THEOREM 1.4. *If* $\mathfrak{G}$ *contains involutions which are not extremal, then a necessary and sufficient condition for an involution* $T \varepsilon \mathfrak{G}$ *to be extremal is that* $\rho_T = \frac{1}{2}\rho$. *A necessary and sufficient condition for every involution in* $\mathfrak{G}$ *to be extremal is that* $\rho_T = \rho$ *for every* T.

Proof. By hypothesis $\dim \mathfrak{X} \geq 6$; therefore the only case in which all involutions are extremal is the symplectic case with $\dim \mathfrak{X} = 6$. Hence, if $\mathfrak{X}$ is symplectic, assume $\dim \mathfrak{X} > 6$. Let T_1, T_2 be two commutative involutions in $\mathfrak{G}$ and set $E_i = \frac{1}{2}T_i$ $(i = 1, 2)$. The E_i are self-adjoint projections and, since E_1, E_2 commute, the following are also self-adjoint projections:

$$P_1 = E_1E_2, \qquad P_2 = E_1(I - E_2),$$

$$P_3 = (I - E_1)E_2, \qquad P_4 = (I - E_1)(I - E_2).$$

Evidently $P_iP_j = 0$ for $i \neq j$ and $I = \Sigma P_i$. Observe that $c(T_1, T_2)$ consists of all involutions $T \varepsilon \mathfrak{G}$ such that T commutes with each P_i. Now let U_i be any minimal involution in $\mathfrak{G}$ such that $P_iU_i = U_iP_i = U_i$. Then $P_jU_i = U_iP_j = 0$, for $i \neq j$, so that $U_i \varepsilon c(T_1, T_2)$. It follows from Lemma 1.3 that every $T \varepsilon c(c(T_1, T_2))$ is of the form $T = \Sigma \delta_i P_i$, where $\delta_i = 0$ or 2. Conversely, every T of this form is in $c(c(T_1, T_2))$ provided only that $T \varepsilon \mathfrak{G}$. If none of the P_i is zero, it follows by direct calculation that $\rho(T_1, T_2) \leq 16$ or 8 according as $2I \varepsilon \mathfrak{G}$ or $2I \notin \mathfrak{G}$. On the other hand, if one of the P_i is zero, then $\rho(T_1, T_2) \leq 8$ or 4 according as $2I \varepsilon \mathfrak{G}$ or $2I \notin \mathfrak{G}$.

Next let T_1 be an arbitrary involution in $\mathfrak{G}$ with subspace $\mathfrak{M}$, where $\mathfrak{M} \neq (0)$ and $\mathfrak{M}^{\perp} \neq (0)$. Let $\mathfrak{M}_1$ be a minimal non-isotropic subspace of $\mathfrak{M}$ and $\mathfrak{M}_2$ a minimal non-isotropic subspace of $\mathfrak{M}^{\perp}$. Also let U_i be a minimal involution with subspace $\mathfrak{M}_i$ $(i = 1, 2)$ and set $T_2 = U_1 \circ U_2$. Evidently U_1, U_2 commute with one another and also with T_1. Therefore T_2 is an involution in $\mathfrak{G}$ which commutes with T_1. With T_2 chosen in this way, at most one of the associated projections P_i is zero and this happens if, and only if, T_1 is extremal. It follows that $\rho(T_1, T_2) = \rho$ if, and only if, T_1 is not extremal and $\rho(T_1, T_2) = \frac{1}{2}\rho$ if, and only if, T_1 is extremal.[9]

The last statement of the theorem follows from the above arguments plus the fact that every involution in $\mathcal{G}$ is extremal if, and only if, $\mathfrak{X}$ is symplectic of dimension six.

In the remainder of this paper we shall be concerned with two self-dual spaces $\mathfrak{X}$ and $\mathfrak{Y}$ of the same type (that is, either both symplectic or both unitary) over division rings $\mathcal{D}$ and $\mathcal{E}$ respectively. The same notations for the involutions and scalar products will be used for both spaces. It will be assumed throughout that the dimensions of $\mathfrak{X}$ and $\mathfrak{Y}$ are at least six, that neither $\mathcal{D}$ nor $\mathcal{E}$ has characteristic two and, in the unitary case, that both $\mathcal{D}$ and $\mathcal{E}$ are different from F_3.

Let $\mathcal{G}$ and $\mathcal{H}$ be subgroups of $\mathcal{U}(\mathfrak{X})$ and $\mathcal{U}(\mathfrak{Y})$ respectively, each of which contains all of the minimal involutions. We assume given a group isomorphism $G \rightarrow \mathfrak{g}(G)$ of $\mathcal{G}$ onto $\mathcal{H}$. In view of Lemma 1. 3, $\mathcal{H}$ will contain $2I$ if, and only if, $\mathcal{G}$ contains $2I$. Furthermore, if $2I \varepsilon \mathcal{G}$, then $\mathfrak{g}(2I) = 2I$. As a consequence of Theorem 1. 4, we have that $\mathfrak{g}(G)$ is an extremal involution in $\mathcal{H}$ if, and only if, G is an extremal involution in $\mathcal{G}$. The purpose of the next three sections is to obtain a representation of the group isomorphism $\mathfrak{g}$. The next part of our discussion has to deal separately with the unitary and symplectic cases.

2. The unitary case.

Throughout this section, both $\mathfrak{X}$ and $\mathfrak{Y}$ are assumed to be unitary. We prove first two lemmas for unitary spaces.

LEMMA 2. 1.[10] *Let u, v_i be isotropic vectors in $\mathfrak{X}$ such that $(u, v_i) = 1$, for $i = 1, \cdots, k$, where $k + 2 < \dim \mathfrak{X}$. Then there exists an isotropic vector w linearly independent of $u, v_i, \cdots, v_k$ and such that $(u, w) = 1$.*

Proof. Since $\dim (\mathcal{D}w)^{\perp} = \dim \mathfrak{X} - 1 > k + 1$, there exists $r \varepsilon (\mathcal{D}u)^{\perp}$ linearly independent of $u, v_1, \cdots, v_k$. Set $w = \frac{1}{2}(s, s)u + s$, where $s = v_1 + r$. Then $(w, w) = 0$, $(u, w) = 1$ and w is linearly independent of $u, v_1, \cdots, v_k$.

LEMMA 2. 2. *Let $u, v_1, \cdots, v_k$ be vectors in $\mathfrak{X}$ such that u is linearly*

---

[9] Observe that in this case $2P_1 = U_1$, $2P_2 = T_1 \circ U_1$, $2P_3 = U_2$, and $2P_4 = (2I) \circ T_1 \circ U_2$. Also $\Sigma \delta_i P_i = (\delta_1 P_1) \circ (\delta_2 P_2) \circ (\delta_3 P_3) \circ (\delta_4 P_4)$. Therefore, if we set

$$V = (\epsilon_1 U_1) \circ (\epsilon_2 (T_1 \circ U_2)) \circ (\epsilon_3 U_2) \text{ and } W = V \circ T_1 \circ U_2,$$

where $\epsilon_i = \frac{1}{2}\delta_i$, then $V, W \varepsilon \mathcal{G}$ and $\Sigma \delta_i P_i = V$ or $2I \circ W$ according as $\epsilon_4 = 0$ or 1.

[10] This lemma, and the lemma which follows, are needed below for the case $k = 3$. Hence the restriction $\dim \mathfrak{X} \geqq 6$ in the unitary case.

independent of $v_1, \cdots, v_k$, *where* $0 \leq 2k \leq \dim \mathfrak{X}$ *and at least one of the vectors* $v_1, \cdots, v_k$ *is non-isotropic. Then there exists a non-isotropic vector* w *such that* $(u, w) = 1$ *and* $(v_i, w) = 0$ *for* $i = 1, \cdots, k$.

Proof. Choose $r \varepsilon \mathfrak{X}$ such that $(u, r) = 1$ and $(v_i, r) = 0$ for $i = 1, \cdots, k$. If $(r, r) \neq 0$, we can take $w = r$. Therefore assume $(r, r) = 0$. Since at least one of the vectors v_i is non-isotropic, the subspace $\mathfrak{D} v_1 + \cdots + \mathfrak{D} v_k$ is not contained in $(\mathfrak{D} v_1 + \cdots + \mathfrak{D} v_k)^{\perp}$. Hence the intersection of these subspaces has dimension at most $k - 1$. On the other hand, since $2k \leq \dim \mathfrak{X}$, we have $\dim (\mathfrak{D} v_1 + \cdots + \mathfrak{D} v_k)^{\perp} \geqq \dim \mathfrak{X} - k \geqq k$. It follows that there exists a vector s linearly independent of $v_1, \cdots, v_k$ such that $(v_i, s) = 0$ for $i = 1, \cdots, k$. Now, if r is linearly dependent on $v_1, \cdots, v_k$, consider $r' = r - (s, u) r + s$. Then r' is linearly independent of $v_1, \cdots, v_k$, $(u, r') = 1$ and $(v_i, r') = 0$ for $i = 1, \cdots, k$. Therefore we may as well assume r to be linearly independent of $v_1, \cdots, v_k$. It is easy to see that r must also be linearly independent of $u, v_1, \cdots, v_k$. Next choose $t \varepsilon \mathfrak{X}$ such that $(r, t) = 1$, $(u, t) = 0$ and $(v_i, t) = 0$ for $i = 1, \cdots, k$. Define $w = r + \lambda t$, where λ is to be determined. Note first that $(v_i, w) = 0$, for $i = 1, \cdots, k$, and $(u, w) = 1$ independently of λ. Furthermore $(w, w) = \lambda + \lambda^* + \lambda (t, t) \lambda^*$. If $(t, t) = 0$, take $\lambda = 1$ to obtain $(w, w) = 2$. If $(t, t) \neq 0$, take $\lambda = -(t, t)^{-1}$ to obtain $(w, w) = -(t, t)^{-1}$. In either case w is non-isotropic and the proof is complete.

LEMMA 2. 3. *Let* T *be an arbitrary two-dimensional involution in* $\mathfrak{U}(\mathfrak{X})$ *with* $\mathfrak{M}$ *as its subspace. If* u *is any non-isotropic vector in* $\mathfrak{M}$, *then* T *can be written in the form* $T = U \circ V$, *where* U *and* V *are minimal involutions in* $\mathfrak{U}(\mathfrak{X})$, $UV = VU = 0$ *and* $\mathfrak{X} U = \mathfrak{D} u$.

Proof. There exists a non-isotropic $v \varepsilon \mathfrak{M}$ such that $(u, v) = 0$. Define $xU = 2(x, u)(u, u)^{-1} u$ and $xV = 2(x, v)(v, v)^{-1} v$. Then U, V are minimal involutions in $\mathfrak{U}(\mathfrak{X})$. Also $UV = VU = 0$ so that $U \circ V = U + V$. Since $U \circ V$ is a two-dimensional involution in $\mathfrak{U}(\mathfrak{X})$ with $\mathfrak{D} u + \mathfrak{D} v = \mathfrak{M}$ as its subspace. It follows that $T = U \circ V$.

LEMMA 2. 4. *Let* T *be a two-dimensional involution in* $\mathfrak{G}$. *Then either* $\mathfrak{g}(T)$ *or* $(2I) \circ \mathfrak{g}(T)$ *is a two-dimensional involution in* $\mathfrak{H}$.

Proof. Write T in the form $T = U \circ V$, where U and V are minimal involutions. Then $\mathfrak{g}(T) = \mathfrak{g}(U) \circ \mathfrak{g}(V)$. Since $\mathfrak{g}(U)$ and $\mathfrak{g}(V)$ are extremal, it follows that either $\mathfrak{g}(T)$ or $(2I) \circ \mathfrak{g}(T)$ is two-dimensional.

Let T_1, T_2 be two-dimensional involutions in $\mathfrak{U}(\mathfrak{X})$ with subspaces $\mathfrak{M}_1$,

$\mathfrak{M}_2$ respectively. The involutions T_1, T_2 are said to *intersect* provided $T_1 \neq T_2$ and $\mathfrak{M}_1 \cap \mathfrak{M}_2 \neq (0)$. Observe that T_1, T_2 intersect if, and only if, $\dim (\mathfrak{M}_1 + \mathfrak{M}_2) = 3$.

LEMMA 2. 5. *Let T_1 and T_2 be two-dimensional involutions in $\mathfrak{G}$. Then the associated two-dimensional involutions in $\mathfrak{H}$ intersect if, and only if, T_1 and T_2 intersect.*

Proof. Let $T_i = U_i \circ V_i$, where U_i, V_i are minimal involutions. It will be sufficient, because of the symmetry, to prove that intersection of T_1 and T_2 implies intersection of the associated two-dimensional involutions in $\mathfrak{H}$. If T_1 and T_2 are assumed to intersect, there is clearly no loss in assuming $\mathfrak{X}V_2$ contained in $\mathfrak{X}U_1 + \mathfrak{X}U_2 + \mathfrak{X}V_1$. Denote by $\mathfrak{N}_i$ the two-dimensional subspace associated with $\mathfrak{g}(T_i)$. Evidently $\mathfrak{N}_1 \neq \mathfrak{N}_2$; otherwise either $\mathfrak{g}(T_1) = \mathfrak{g}(T_2)$ or $\mathfrak{g}(T_1) = (2I) \circ \mathfrak{g}(T_2)$ which implies either $T_1 = T_2$ or $T_1 = (2I) \circ T_2$, neither of which is possible. Therefore we have only to prove $\dim (\mathfrak{N}_1 + \mathfrak{N}_2) = 3$. Choose vectors $r_i, s_i \in \mathfrak{Y}$ such that $\mathcal{E}r_i$ and $\mathcal{E}s_i$ are the one-dimensional subspaces of $\mathfrak{Y}$ associated with $\mathfrak{g}(U_i)$ and $\mathfrak{g}(V_i)$ respectively. Then $\mathfrak{N}_i = \mathcal{E}r_i + \mathcal{E}s_i$; hence, if $\dim (\mathfrak{N}_1 + \mathfrak{N}_2) \neq 3$, then r_1, s_1, r_2, s_2 are linearly independent non-isotropic vectors in $\mathfrak{Y}$. By Lemma 2.2, there exists a non-isotropic vector $t \in \mathfrak{Y}$ such that $(r_1, t) = (s_1, t) = (r_2, t) = 0$ while $(s_2, t) = 1$. Let W be a minimal involution in $\mathfrak{G}$ such that $\mathcal{E}t$ is the one-dimensional subspace associated with the extremal involution $\mathfrak{g}(W)$. Evidently $\mathfrak{g}(W)$ commutes with and is different from each of the involutions $\mathfrak{g}(U_1)$, $\mathfrak{g}(V_1)$, $\mathfrak{g}(U_2)$. Therefore W commutes with and is different from each of the involutions U_1, V_1, U_2. It follows by the Corollary to Lemma 1. 2, that $\mathfrak{X}W$ is orthogonal to each of the subspaces $\mathfrak{X}U_1$, $\mathfrak{X}V_1$, $\mathfrak{X}U_2$. Since, by hypothesis, $\mathfrak{X}V_2$ is contained in $\mathfrak{X}U_1 + \mathfrak{X}V_1 + \mathfrak{X}U_2$, it is also true that $\mathfrak{X}W$ is orthogonal to $\mathfrak{X}V_2$. Therefore $W \neq V_2$ and W, V_2 commute. It follows that $\mathfrak{g}(W) \neq \mathfrak{g}(V_2)$ and $\mathfrak{g}(W)$, $\mathfrak{g}(V_2)$ commute. But commutativity of $\mathfrak{g}(W)$, $\mathfrak{g}(V_2)$ implies that either $\mathcal{E}t = \mathcal{E}s_2$ or $(\mathcal{E}t) \perp (\mathcal{E}s_2)$. The second possibility is ruled out since $(s_2, t) = 1$. Hence $\mathcal{E}t = \mathcal{E}s_2$. Since $\mathfrak{g}(W) \neq \mathfrak{g}(V_2)$, this implies that $\mathfrak{g}(W) = (2I) \circ \mathfrak{g}(V_2) = \mathfrak{g}((2I) \circ V_2)$. Therefore $W = (2I) \circ V_2$. But this is impossible since both W and V_2 are minimal. It follows that $\dim (\mathfrak{N}_1 + \mathfrak{N}_2) = 3$ and the proof is complete.

The group isomorphism $G \rightarrow \mathfrak{g}(G)$ will now be used to construct a one-to-one mapping of the one-dimensional subspaces of $\mathfrak{X}$ onto the one-dimensional subspaces of $\mathfrak{Y}$.

THEOREM 2. 6. *There exists a one-to-one mapping F of the one-*

dimensional subspaces of $\mathfrak{X}$ *onto the one-dimensional subspaces of* $\mathfrak{Y}$ *with the following properties*:

(i) *If* U *is an extremal involution in* $\mathfrak{G}$ *with associated one-dimensional subspace* $\mathfrak{D}u$, *then* $(\mathfrak{D}u)F$ *is the one-dimensional subspace associated with* $\mathfrak{g}(U)$.

(ii) $(\mathfrak{D}u)F \perp (\mathfrak{D}v)F$ *if, and only if,* $\mathfrak{D}u \perp \mathfrak{D}v$.

(iii) $(\mathfrak{D}u_3)F \subseteq (\mathfrak{D}u_1)F + (\mathfrak{D}w_2)F$ *if, and only if,* $\mathfrak{D}u_3 \subseteq \mathfrak{D}u_1 + \mathfrak{D}u_2$.

Proof. Consider first a non-isotropic one-dimensional subspace $\mathfrak{D}u$ in $\mathfrak{X}$ and let U be the minimal involution in $\mathfrak{G}$ which has $\mathfrak{D}u$ as its subspace. Then $\mathfrak{g}(U)$ is an extremal involution and therefore either the subspace $\mathfrak{N}$ of $\mathfrak{g}(U)$ or its orthogonal complement $\mathfrak{N}^{\perp}$ is a non-isotropic one-dimensional subspace of $\mathfrak{Y}$. In the first case we define $(\mathfrak{D}u)F = \mathfrak{N}$ and in the second $(\mathfrak{D}u)F = \mathfrak{N}^{\perp}$. This evidently establishes a one-to-one correspondence between the non-isotropic one dimensional subspaces of $\mathfrak{X}$ and $\mathfrak{Y}$.

If $\mathfrak{D}u$ is isotropic, then, by Lemma 2.1, it is possible to choose two isotropic vectors v_1, v_2 such that u, v_1, v_2 are linearly independent and $(u, v_1) = (u, v_2) = 1$. Then $\mathfrak{M}_i = \mathfrak{D}u + \mathfrak{D}v_i$ $(i = 1, 2)$ are non-isotropic, two-dimensional subspaces of $\mathfrak{X}$ which intersect in $\mathfrak{D}u$. Let T_i be the involution with $\mathfrak{M}_i$ as its subspace. Then T_1, T_2 intersect and, by Lemma 2.5, the two-dimensional subspaces associated with $\mathfrak{g}(T_1)$ and $\mathfrak{g}(T_2)$ intersect in a one-dimensional subspace of $\mathfrak{Y}$ which we define to be $(\mathfrak{D}u)F$. Observe that $(\mathfrak{D}u)F$ is isotropic; otherwise it would be possible to find extremal involutions R, S_i in $\mathfrak{H}$ such that $\mathfrak{g}(T_i) = R \circ S_i$ and $\mathfrak{Y}R = (\mathfrak{D}w)F$. But then $T_i = U \circ V_i$, where U, V_i are minimal involutions in $\mathfrak{G}$, and this implies that $\mathfrak{D}u = \mathfrak{M}_1 \cap \mathfrak{M}_2 = \mathfrak{X}U$, contradicting the assumption that $\mathfrak{D}u$ be isotropic.

We now have $(\mathfrak{D}u)F$ defined for all $\mathfrak{D}u$. The next step is to prove that F is single-valued. This is obvious if $\mathfrak{D}u$ is non-isotropic. Therefore let $\mathfrak{D}u$ be isotropic and let T_1, T_2 be the intersecting two-dimensional involutions used in the definition of $(\mathfrak{D}u)F$. The problem is to show that $(\mathfrak{D}u)F$ is independent of the choice of T_1, T_2. This is equivalent to showing that, if T_3 is any other two-dimensional involution in $\mathfrak{G}$ for which $\mathfrak{D}u \subseteq \mathfrak{X}T_3$, then the two-dimensional subspace associated with $\mathfrak{g}(T_3)$ contains $(\mathfrak{D}u)F$. Denote the two-dimensional subspace of T_i by $\mathfrak{M}_i$ and the two-dimensional subspace associated with $\mathfrak{g}(T_i)$ by $\mathfrak{N}_i$. Consider first the case in which $\mathfrak{M}_3 \not\subseteq \mathfrak{M}_1 + \mathfrak{M}_2$; then $\dim(\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3) = 4$. We have, by definition, that $(\mathfrak{D}u)F \subseteq \mathfrak{N}_1 \cap \mathfrak{N}_2$ and, by Lemma 2.5, the spaces $\mathfrak{N}_1$, $\mathfrak{N}_2$, $\mathfrak{N}_3$ intersect by pairs. Therefore it will be sufficient to prove that $\dim(\mathfrak{N}_1 + \mathfrak{N}_2 + \mathfrak{N}_3) = 4$ since this will imply $\dim(\mathfrak{N}_1 \cap \mathfrak{N}_2 \cap \mathfrak{N}_3) = 1$, which in turn will imply $(\mathfrak{D}u)F \subseteq \mathfrak{N}_3$. Let $T_i = R_i \circ S_i$, where R_i and S_i are minimal involutions in $\mathfrak{G}$. Observe that $\mathfrak{M}_i = \mathfrak{X}R_i + \mathfrak{X}S_i$ and also that $\mathfrak{N}_i = (\mathfrak{X}R_i)F + (\mathfrak{X}S_i)F$. Since $\dim(\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3) = 4$, we have $\dim(\mathfrak{X}R_1 + \mathfrak{X}S_1 + \mathfrak{X}R_2 + \mathfrak{X}R_3) = 4$. Hence, by Lemma 2. 2, there exists a one-dimensional non-isotropic subspace $\mathfrak{D}w$ in $\mathfrak{X}$ which is orthogonal to $\mathfrak{X}R_1$, $\mathfrak{X}S_1$ and $\mathfrak{X}R_2$ but not to $\mathfrak{X}R_3$. Then, as in the proof of Lemma 2. 5, $(\mathfrak{D}w)F$ is orthogonal to $(\mathfrak{X}R_1)F$, $(\mathfrak{X}S_1)F$ and $(\mathfrak{X}R_2)F$ but not to $(\mathfrak{X}R_3)F$. This shows that $(\mathfrak{X}R_3)F$ is not contained in $(\mathfrak{X}R_1)F + (\mathfrak{X}S_1)F + (\mathfrak{X}R_2)F$. Similarly none of the subspaces $(\mathfrak{X}R_1)F$, $(\mathfrak{X}S_1)F$, $(\mathfrak{X}R_2)F$, $(\mathfrak{X}R_3)F$ is contained in the union of the remaining three. In other words these subspaces are linearly independent. Therefore it follows that $\dim(\mathfrak{N}_1 + \mathfrak{N}_2 + \mathfrak{N}_3) = 4$ and hence $(\mathfrak{D}u)F \subseteq \mathfrak{N}_3$.

Now assume that $\mathfrak{M}_3 \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. Let v_i be isotropic vectors such that $(u, v_i) = 1$ and $\mathfrak{M}_i = \mathfrak{D}u + \mathfrak{D}v_i$ $(i = 1, 2, 3)$. By Lemma 2. 1, there exists an isotropic vector v_4 linearly independent of u, v_1, v_2, v_3 and such that $(u, v_4) = 1$. Set $\mathfrak{M}_4 = \mathfrak{D}u + \mathfrak{D}v_4$. Then $\mathfrak{M}_4 \not\subseteq \mathfrak{M}_1 + \mathfrak{M}_2$ and $\mathfrak{M}_3 \not\subseteq \mathfrak{M}_1 + \mathfrak{M}_4$. Let T_4 be the two-dimensional involution in $\mathfrak{G}$ with subspace $\mathfrak{M}_4$ and denote by $\mathfrak{N}_4$ the two-dimensional subspace associated with $\mathfrak{g}(T_4)$. Then, by the above argument, $(\mathfrak{D}u)F \subseteq \mathfrak{N}_4$ and therefore $(\mathfrak{D}u)F = \mathfrak{N}_1 \cap \mathfrak{N}_4$. Another application of the same argument, using T_4 instead of T_2 gives $(\mathfrak{D}u)F \subseteq \mathfrak{N}_3$ and completes the proof that F is single-valued. By symmetry, F is a one-to-one mapping of the one-dimensional subspaces of $\mathfrak{X}$ *onto* the one-dimensional subspaces of $\mathfrak{Y}$.

Property (i) is immediate from the definition of F. Property (ii), for non-isotropic subspaces, has already been observed in the proof of Lemma 2. 5. Next let $\mathfrak{D}u$ be isotropic, $\mathfrak{D}v$ non-isotropic, and $\mathfrak{D}u \perp \mathfrak{D}v$. Choose a vector $r \in \mathfrak{X}$ such that $(u, r) = 1$, $(v, r) = 0$ and set $w = -\frac{1}{2}(r, r)u + r$. Then $(u, w) = 1$ and $(w, w) = 0$. Evidently $u + w$, $u - w$ are non-isotropic and orthogonal to v. Therefore $(\mathfrak{D}v)F$ is orthogonal to $(\mathfrak{D}(u + w))F$ and $(\mathfrak{D}(u - w))F$. Since $\mathfrak{D}u \subseteq \mathfrak{D}(u + w) + \mathfrak{D}(u - w)$, the proof that F is single-valued gives $(\mathfrak{D}u)F \subseteq (\mathfrak{D}(u + w))F + (\mathfrak{D}(u - w))F$. Therefore $(\mathfrak{D}u)F \perp (\mathfrak{D}v)F$. Now let both $\mathfrak{D}u$ and $\mathfrak{D}v$ be isotropic with $\mathfrak{D}u \perp \mathfrak{D}v$. Choose w as before; then $\mathfrak{D}v$ is orthogonal to $\mathfrak{D}(u + w)$ and $\mathfrak{D}(u - w)$. Hence, by the result just obtained, $(\mathfrak{D}v)F$ is orthogonal to $(\mathfrak{D}(u + w))F$ and $(\mathfrak{D}(u - w))F$. Therefore again $(\mathfrak{D}u)F \perp (\mathfrak{D}v)F$. Property (ii) now follows by a symmetry argument.

In order to prove (iii), it will be sufficient to prove that $\mathfrak{D}u_3 \subseteq \mathfrak{D}u_1 + \mathfrak{D}u_2$ implies $(\mathfrak{D}u_3)F \subseteq (\mathfrak{D}u_1)F + (\mathfrak{D}u_2)F$, where $\mathfrak{D}u_3 \neq \mathfrak{D}u_1$, $\mathfrak{D}u_3 \neq \mathfrak{D}u_2$ and hence $\mathfrak{D}u_1 \neq \mathfrak{D}u_2$. Suppose $(\mathfrak{D}u_3)F \not\subseteq (\mathfrak{D}u_1)F + (\mathfrak{D}u_2)F$. Then there exists $\mathfrak{D}u_4$ such that $(\mathfrak{D}u_4)F$ is orthogonal to $(\mathfrak{D}u_1)F$ and $(\mathfrak{D}u_2)F$ but not to $(\mathfrak{D}u_3)F$. By (ii), this implies that $\mathfrak{D}u_4$ is orthogonal to $\mathfrak{D}u_1$ and $\mathfrak{D}u_2$ but not to $\mathfrak{D}u_3$. This contradicts $\mathfrak{D}u_3 \subseteq \mathfrak{D}u_1 + \mathfrak{D}u_2$ and completes proof of the lemma.

3. The simplectic case. Throughout this section both $\mathfrak{X}$ and $\mathfrak{Y}$ are assumed to be symplectic. The objective here is to obtain the analogue of Theorem 1. 6 for this case. Recall that all of the minimal involutions in the symplectic case are two-dimensional.

LEMMA 3. 1. *Let T_1, T_2 be non-commutative minimal involutions in $\mathfrak{G}$ with subspaces $\mathfrak{M}_1$, $\mathfrak{M}_2$ respectively. A necessary and sufficient condition for an involution T in $\mathfrak{G}$ to belong to $c(T_1, T_2)$ is that $\mathfrak{M}_1 + \mathfrak{M}_2$ be contained in either the subspace of T or its orthogonal complement.*

Proof. The sufficiency is immediate from the first part of Lemma 1. 2. On the other hand, if $T \,\varepsilon\, c(T_1, T_2)$ and has $\mathfrak{M}$ as its subspace, then by the second part of Lemma 1. 2, either $\mathfrak{M}_i \subseteq \mathfrak{M}$ or $\mathfrak{M}_i \subseteq \mathfrak{M}^\perp$. We have only to show that the same case occurs for both $i = 1, 2$. Suppose $\mathfrak{M}_1 \subseteq \mathfrak{M}$ and $\mathfrak{M}_2 \subseteq \mathfrak{M}^\perp$. Since $\mathfrak{M}_1 \subseteq \mathfrak{M}$, we have $\mathfrak{M}^\perp \subseteq \mathfrak{M}_1^\perp$. Therefore $\mathfrak{M}_2 \subseteq \mathfrak{M}_1^\perp$ and hence $\mathfrak{M}_1 \subseteq \mathfrak{M}_2^\perp$. This implies $T_1T_2 = T_2T_1 = 0$ which is contrary to the hypothesis that T_1, T_2 are non-commutative. It follows that either $\mathfrak{M}_1 + \mathfrak{M}_2 \subseteq \mathfrak{M}$ or $\mathfrak{M}_1 + \mathfrak{M}_2 \subseteq \mathfrak{M}^\perp$.

LEMMA 3. 2. *Let T_1, T_2 be non-commutative extremal involutions in $\mathfrak{G}$. A sufficient condition for $T \,\varepsilon\, c(c(T_1, T_2))$ is that either the subspace of T or its orthogonal complement be contained in the union of the two-dimensional subspaces associated with T_1, T_2. The condition is also necessary if* $\dim \mathfrak{X} > 6$ *or if T_1, T_2 intersect.*

Proof. There is no loss in assuming T_i minimal with subspace $\mathfrak{M}_i$ $(i = 1, 2)$. Let T be an involution in $\mathfrak{G}$ which satisfies the given condition and let $T' \,\varepsilon\, c(T_1, T_2)$. By Lemma 3. 1, $\mathfrak{M}_1 + \mathfrak{M}_2$ is contained either in the subspace of T' or its orthogonal complement. Thus either the subspace of T or its orthogonal complement is contained either in the subspace of T' or its orthogonal complement. By Lemma 1. 2, this implies that T, T' commute. Therefore $T \,\varepsilon\, c(c(T_1, T_2))$ and the sufficiency is proved.

Now assume $\dim \mathfrak{X} > 6$ and let $\mathfrak{N}$ denote any complement of $\mathfrak{M}_1^\perp \cap \mathfrak{M}_2^\perp$ in $\mathfrak{M}_1^\perp$. Since $\mathfrak{M}_2$ is a complement of $\mathfrak{M}_2^\perp$ in $\mathfrak{X}$ and $\dim \mathfrak{M}_2 = 2$, it follows that $\dim \mathfrak{N} \leq 2$. We have $\mathfrak{X} = \mathfrak{M}_1 \oplus (\mathfrak{M}_1^\perp \cap \mathfrak{M}_2^\perp) \oplus \mathfrak{N}$. Hence

$\dim \mathfrak{X} - 2 \leq \dim (\mathfrak{M}_1 \oplus (\mathfrak{M}_1^{\perp} \cap \mathfrak{M}_2^{\perp}))$. Suppose $\mathfrak{M}_1^{\perp} \cap \mathfrak{M}_2^{\perp} \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. Then $\dim (\mathfrak{M}_1 \oplus (\mathfrak{M}_1^{\perp} \cap \mathfrak{M}_2^{\perp})) \leq \dim (\mathfrak{M}_1 + \mathfrak{M}_2) \leq 4$. This implies $\dim \mathfrak{X} \leq 6$ contrary to hypothesis. Therefore $\mathfrak{M}_1^{\perp} \cap \mathfrak{M}_2^{\perp} \not\subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. Now take $T \varepsilon c(c(T_1, T_2))$ and let $\mathfrak{M}$ be the subspace of T. Also let y, z be arbitrary elements of $\mathfrak{X}$ such that $y, z \not\varepsilon \mathfrak{M}_1 + \mathfrak{M}_2$ and $z \varepsilon \mathfrak{M}_1^{\perp} \cap \mathfrak{M}_2^{\perp}$. Since $\mathfrak{M}_1 + \mathfrak{M}_2$ is finite-dimensional and $y, z \varepsilon \mathfrak{M}_1 + \mathfrak{M}_2$, there exists $s \varepsilon \mathfrak{X}$ such that $(\mathfrak{M}_1 + \mathfrak{M}_2, s) = (0)$, $(z, s) = 1$ and $(y, s) \neq 0$. Define Z by $xZ = 2(x, s)z - 2(x, z)s$. Then $Z \varepsilon c(T_1, T_2)$. Therefore Z, T commute. By Lemma 1. 2, either $\mathfrak{M}$ or $\mathfrak{M}^{\perp}$ must contain both s and z. Suppose $s, z \varepsilon \mathfrak{M}$. Then $(\mathfrak{M}^{\perp}, s) = (\mathfrak{M}^{\perp}, z) = (0)$. Therefore $y \not\varepsilon \mathfrak{M}^{\perp}$. Holding z fixed and letting y vary over all vectors not in $\mathfrak{M}_1 + \mathfrak{M}_2$, we see that $\mathfrak{M}^{\perp} \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. Similarly, if $z \varepsilon \mathfrak{M}^{\perp}$, then we obtain $\mathfrak{M} \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. This proves the necessity when $\dim \mathfrak{X} > 6$. Now the condition $\dim \mathfrak{X} > 6$ was used only to ensure that $\mathfrak{M}_1^{\perp} \cap \mathfrak{M}_2^{\perp} \not\subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. If T_1, T_2 intersect, then $\dim (\mathfrak{M}_1 + \mathfrak{M}_2) = 3$ so that $\mathfrak{M}_1^{\perp} \cap \mathfrak{M}_2^{\perp}$ cannot be contained in $\mathfrak{M}_1 + \mathfrak{M}_2$ even when $\dim \mathfrak{X} = 6$. Therefore the above argument applies and the proof is complete.

LEMMA 3. 3. *Let T_i $(i = 1, 2, 3)$ be non-commutative minimal involutions in $\mathfrak{G}$ which intersect by pairs and let $\mathfrak{M}_i$ be the subspace of T_i. Then always* $\dim (\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3) = 3$ *or* 4 *and is equal to* 3 *if, and only if,* $T_3 \varepsilon c(c(T_1, T_2))$.

Proof. That $\dim (\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3) = 3$ or 4 is obvious. Moreover, if $\dim (\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3) = 3$, then $\mathfrak{M}_3 \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$ and $T_3 \varepsilon c(c(T_1, T_2))$ by Lemma 3. 2. Conversely, if $T_3 \varepsilon c(c(T_1, T_2))$, then, again by Lemma 3. 2. $\mathfrak{M}_3 \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$ and hence $\dim (\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3) = 3$.

THEOREM 3. 4. *Let T_1, T_2 be non-commutative extremal involutions in $\mathfrak{G}$. Then a necessary and sufficient condition for T_1, T_2 to intersect is that $c(c(S_1, S_2)) = c(c(T_1, T_2))$ for every pair S_1, S_2 of non-commutative extremal involutions in $c(c(T_1, T_2))$.*

Proof. There is obviously no loss in assuming the T_i to be minimal with subspaces $\mathfrak{M}_i$ and written in the form $xT_i = 2(x, u_i)t_i - 2(x, t_i)u_i$, $(t_i, u_i) = 1$ $(i = 1, 2)$. Assume first that T_1, T_2 do not intersect so that $\dim (\mathfrak{M}_1 + \mathfrak{M}_2) = 4$. Define $s_1 = t_1 + t_2 - (t_1, u_2)t_2$. Then s_1, t_2, u_2 are linearly independent and $(s_1, u_2) = 1$. Define $xS_1 = 2(x, u_2)s_1 - 2(x, s_1)u_2$ and take $S_2 = T_2$. Then S_1 is an involution in $\mathfrak{G}$. Moreover, if $\mathfrak{N}_i$ is the subspace of S_i, then $\mathfrak{N}_i \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. Therefore $S_i \varepsilon c(c(T_1, T_2))$, by Lemma 3. 2. On the other hand, if $\mathfrak{M}_1 \subseteq \mathfrak{N}_1 + \mathfrak{N}_2$, then $\mathfrak{N}_1 + \mathfrak{N}_2$ contains t_1, u_1, t_2, u_2 and hence coincides with $\mathfrak{M}_1 + \mathfrak{M}_2$. But $\dim (\mathfrak{N}_1 + \mathfrak{N}_2) = 3$ and dim $(\mathfrak{M}_1 + \mathfrak{M}_2) = 4$. Therefore $\mathfrak{M}_1 \not\subseteq \mathfrak{N}_1 + \mathfrak{N}_2$. Since S_1, S_2 intersect, this implies, by Lemma 3. 2, that $T_1 \notin c(c(S_1, S_2))$ and proves the sufficiency.

Now assume that T_1, T_2 intersect so that dim $(\mathfrak{M}_1 + \mathfrak{M}_2) = 3$. Let S_1, S_2 be any pair of non-commutative extremal involutions in $c(c(T_1, T_2))$. We can evidently assume that S_i is minimal with subspace $\mathfrak{N}_i$ $(i = 1, 2)$. Since T_1, T_2 intersect, we have, by Lemma 3. 2, that $\mathfrak{N}_1 + \mathfrak{N}_2 \subseteq \mathfrak{M}_1 + \mathfrak{M}_2$. Since S_1, S_2 do not commute, dim $(\mathfrak{N}_1 + \mathfrak{N}_2) \geq 3$. Therefore $\mathfrak{N}_1 + \mathfrak{N}_2 = \mathfrak{M}_1 + \mathfrak{M}_2$ and, by Lemma 3. 2, it follows that $c(c(S_1, S_2)) = c(c(T_1, T_2))$. This completes the proof.

We are now in a position to use the group isomorphism $G \to \mathfrak{g}(G)$ to construct a one-to-one mapping of the one-dimensional subspaces of $\mathfrak{X}$ onto the one-dimensional subspaces of $\mathfrak{Y}$. Observe first that, if T is an extremal involution in $\mathfrak{G}$ with associated non-isotropic, two-dimensional subspace $\mathfrak{M}$, then $\mathfrak{g}(T)$ is extremal, so determines a non-isotropic, two-dimensional subspace of $\mathfrak{Y}$ which we denote by $\mathfrak{M}\bar{F}$. Evidently $\bar{F}$ is a one-to-one mapping of the non-isotropic two-dimensional subspaces of $\mathfrak{X}$ onto those of $\mathfrak{Y}$. Furthermore, by Theorem 3. 4, $\bar{F}$ preserves the intersection properties of these subspaces.

THEOREM 3. 5. *There exists a one-to-one mapping F of the one-dimensional subspaces of $\mathfrak{X}$ onto the one-dimensional subspaces of $\mathfrak{Y}$ with the properties:*

(i) *If $\mathfrak{M}$ is a non-isotropic two-dimensional subspaces of $\mathfrak{X}$, then $(\mathfrak{D}u)F \subseteq \mathfrak{M}\bar{F}$ if, and only if, $\mathfrak{D}u \subseteq \mathfrak{M}$.*

(ii) *$(\mathfrak{D}u)F \perp (\mathfrak{D}v)F$ if, and only if, $\mathfrak{D}u \perp \mathfrak{D}v$.*

(iii) *$(\mathfrak{D}u_3)F \subseteq (\mathfrak{D}u_1)F + (\mathfrak{D}_2)F$ if, and only if, $\mathfrak{D}u_3 \subseteq \mathfrak{D}u_1 + \mathfrak{D}u_2$.*

Proof. Let $\mathfrak{D}u$ be an arbitrary one-dimensional subspace of $\mathfrak{X}$. Choose $w_1 \varepsilon \mathfrak{X}$ such that $(u, u_1) = 1$. Since dim $(\mathfrak{D}u)^{\perp} \geq 5$, there exists $u'_1 \varepsilon (\mathfrak{D}u)^{\perp}$ such that u, u_1, u'_1 are linearly independent. Define $u_2 = u_1 + u'_1$. Then $(u, u_2) = 1$ and u, u_1, u_2 are linearly independent. Let $\mathfrak{M}_i = \mathfrak{D}u + \mathfrak{D}u_i$. Then $\mathfrak{M}_i$ is a non-isotropic, two-dimensional subspace of $\mathfrak{X}$ and $\mathfrak{M}_1 \cap \mathfrak{M}_2 = \mathfrak{D}u$. It follows that $\mathfrak{M}_1\bar{F}$ and $\mathfrak{M}_2\bar{F}$ intersect in a one-dimensional subspace of $\mathfrak{Y}$ which we denote by $(\mathfrak{D}u)F$. The first problem is to show that $(\mathfrak{D}u)F$ is single-valued; that is, that $(\mathfrak{D}u)F$ is independent of the choice of $\mathfrak{M}_1$, $\mathfrak{M}_2$. This result will be established if we prove that a non-isotropic, two-dimensional subspace $\mathfrak{M}$ in $\mathfrak{X}$ contains $\mathfrak{D}w$ if, and only if, $\mathfrak{M}\bar{F}$ contains $(\mathfrak{D}u)F$. This will also establish property (i). We can obviously assume $\mathfrak{M} \neq \mathfrak{M}_i$ $(i = 1, 2)$ and, because of symmetry, it will be sufficient to prove that $\mathfrak{D}u \subseteq \mathfrak{M}$

implies $(\mathfrak{D}u)F \subseteq \mathfrak{M}\bar{F}$. If dim $(\mathfrak{M} + \mathfrak{M}_1 + \mathfrak{M}_2) = 4$, then it follows from Lemma 3.3 that dim $(\mathfrak{M}\bar{F} + \mathfrak{M}_1\bar{F} + \mathfrak{M}_2\bar{F}) = 4$. Since $\mathfrak{M}\bar{F}$, $\mathfrak{M}_1\bar{F}$, $\mathfrak{M}_2\bar{F}$ intersect by pairs and $\mathfrak{M}_1\bar{F} \cap \mathfrak{M}_2\bar{F} = (\mathfrak{D}u)F$, we obtain $(\mathfrak{D}u)F \subseteq \mathfrak{M}\bar{F}$. Now let dim $(\mathfrak{M} + \mathfrak{M}_1 + \mathfrak{M}_2) = 3$ and choose $r \varepsilon (\mathfrak{D}u)^{\perp}$ such that u, u_1, u_2, r are linearly independent. Define $u_3 = u_1 + r$; then $(u, u_3) = 1$ and u, w_1, u_2, u_3 are linearly independent. If $\mathfrak{M}_3 = \mathfrak{D}u + \mathfrak{D}u_3$, then $\mathfrak{M}_3$ is non-isotropic and dim $(\mathfrak{M} + \mathfrak{M}_2 + \mathfrak{M}_3) =$ dim $(\mathfrak{M} + \mathfrak{M}_1 + \mathfrak{M}_3) = 4$. By the preceding argument, $(\mathfrak{D}u)F \subseteq \mathfrak{M}_3\bar{F}$ and this implies, again by the same argument, that $(\mathfrak{D}u)F \subseteq \mathfrak{M}\bar{F}$.

Next let $\mathfrak{D}u$, $\mathfrak{D}v$ be non-orthogonal one-dimensional subspaces of $\mathfrak{X}$. Then $(u, v) \neq 0$ and the two-dimensional subspace $\mathfrak{M} = \mathfrak{D}u + \mathfrak{D}v$ is non-isotropic. By property (i), we have $\mathfrak{M}\bar{F} = (\mathfrak{D}u)F + (\mathfrak{D}v)F$. Since $\mathfrak{M}\bar{F}$ is non-isotropic, $(\mathfrak{D}u)F$ and $(\mathfrak{D}v)F$ cannot be orthogonal. This result plus symmetry proves (ii).

In order to prove (iii), it will be sufficient to prove that $\mathfrak{D}u_3 \subseteq \mathfrak{D}u_1 + \mathfrak{D}u_2$ implies $(\mathfrak{D}u_3)F \subseteq (\mathfrak{D}u_1)F + (\mathfrak{D}u_2)F$. Suppose on the contrary that $(\mathfrak{D}u_3)F \not\subseteq (\mathfrak{D}u_1)F + (\mathfrak{D}u_2)F$ and choose a non-zero vector $v_i \varepsilon (\mathfrak{D}u_i)F$. Evidently v_1, v_2, v_3 are linearly independent so that there exists $v_4 \varepsilon \mathfrak{Y}$ such that $(v_1, v_4) = (v_2, v_4) = 0$ while $(v_3, v_4) = 1$. Let $\mathfrak{D}w_4$ be chosen so that $(\mathfrak{D}u_4)F = \mathcal{E}v_4$. Then, since $\mathcal{E}v_4$ is orthogonal to $(\mathfrak{D}u_1)F$ and $(\mathfrak{D}u_2)F$ but not to $(\mathfrak{D}u_3)F$, it follows, by property (ii), that $\mathfrak{D}u_4$ is orthogonal to $\mathfrak{D}u_1$ and $\mathfrak{D}u_2$ but not to $\mathfrak{D}u_3$. Since this result is inconsistent with $\mathfrak{D}u_3 \subseteq \mathfrak{D}u_1 + \mathfrak{D}u_2$, the proof is complete.

4. Representation of the group isomorphism. We prove here that the group isomorphism $G \rightarrow \mathfrak{g}(G)$, in both the unitary and symplectic cases, is essentially generated by a linear space isomorphism of $\mathfrak{X}$ onto $\mathfrak{Y}$. The two cases are considered simultaneously.

By a *linear space isomorphism* of $\mathfrak{X}$ onto $\mathfrak{Y}$ we mean a one-to-one semi-linear mapping $x \rightarrow x\Phi$ of $\mathfrak{X}$ onto $\mathfrak{Y}$. In other words, $x \rightarrow x\Phi$ is additive and, for all $\lambda \varepsilon \mathfrak{D}$ and $x \varepsilon \mathfrak{X}$, $(\lambda x)\Phi = \lambda^{\phi}(x\Phi)$, where $\lambda \rightarrow \lambda^{\phi}$ is an isomorphic mapping of the division ring $\mathfrak{D}$ onto $\mathcal{E}$ which is independent of x. The isomorphism Φ is said to *preserve orthongonality* provided $(x\Phi, y\Phi) = 0$ if, and only if, $(x, y) = 0$.

THEOREM 4.1. *Let $x \rightarrow x\Phi$ be a linear space isomorphism of $\mathfrak{X}$ onto $\mathfrak{Y}$ which preserves orthogonality. Then there exists a constant $\alpha \varepsilon \mathcal{E}$ such that $\alpha^* = \alpha$ and*

(i) $\lambda^{\phi *} = \alpha^{-1}\lambda^{*\phi}\alpha$, *for all* $\lambda \varepsilon \mathfrak{D}$,

(ii) $(x\Phi, y\Phi) = (x, y)^{\phi}\alpha$, *for all* $x, y \varepsilon \mathfrak{X}$,

(iii) $G \varepsilon \mathcal{U}(\mathfrak{X})$ *implies* $\Phi^{-1}G\Phi \varepsilon \mathcal{U}(\mathfrak{Y})$.

Proof. Let x, u be arbitrary elements of $\mathfrak{X}$ with $u \neq 0$ and choose v such that $(v, u) = 1$. If $x' = x - (x, u)\mathrm{v}$, then $(x', u) = 0$. Therefore $(x'\Phi, u\Phi) = 0$, and hence $(x\Phi, u\Phi) = (x, u)^{\phi}(v\Phi, u\Phi)$. Define $\alpha_u = (v\Phi, u\Phi)$ and observe that α_u depends on u but not on x. Now let w be any other non-zero element of $\mathfrak{X}$. By additivity we have $(x, u)^{\phi}(\alpha_u - \alpha_{u+w}) = (x, w)^{\phi}(\alpha_{u+w} - \alpha_w)$, for all x. If u, w are linearly independent, it is immediate that $\alpha_u = \alpha_{u+w} = \alpha_w$ and hence that $\alpha_u = \alpha_w$. If w, w are linearly dependent, choose r linearly independent of u, w. Then $\alpha_u = \alpha_r$ and $\alpha_r = \alpha_w$ so that $\alpha_u = \alpha_w$ in this case as well. It follows that α_u is a constant which we denote by α. Thus $(x\Phi, y\Phi) = (x, y)^{\phi}\alpha$ for all x, y. This proves (ii). If $(v, u) = 1$, then $\alpha = (v\Phi, u\Phi) = \epsilon(u\Phi, v\Phi)^* = \epsilon\alpha^*(u, v)^{\phi *} = \alpha^*(v, u)^{\phi *} = \alpha^*$. That is, $\alpha = \alpha^*$. Moreover, for arbitrary $\lambda \varepsilon \mathfrak{D}$, $(v\Phi, (\lambda u)\Phi) = (v, \lambda u)^{\phi}\alpha = (v, u)^{\phi}\lambda^{*\phi}\alpha = \lambda^{*\phi}\alpha$. Also $(v\Phi, (\lambda u)\Phi) = (v\Phi, \lambda^{\phi}u\Phi) = (v\Phi, u\Phi)\lambda^{\phi *} = \alpha\lambda^{\phi *}$. Therefore $\lambda^{\phi *} = \alpha^{-1}\lambda^{*\phi}\alpha$, which is property (i). If $G \varepsilon \mathcal{U}(\mathfrak{X})$, then, for $x, y \varepsilon \mathfrak{Y}$,

$$\begin{aligned}(x\Phi^{-1}G\Phi, y\Phi^{-1}G\Phi) &= (x\Phi^{-1}G, y\Phi^{-1}G)^{\phi}\alpha \\ &= (x\Phi^{-1}G, y\Phi^{-1})^{\phi}\alpha + (x\Phi^{-1}, y\Phi^{-1}G)^{\phi}\alpha \\ &= (x\Phi^{-1}G\Phi, y) + (x, y\Phi^{-1}G\Phi).\end{aligned}$$

Therefore $\Phi^{-1}G\Phi \varepsilon \mathcal{U}(\mathfrak{Y})$. This completes the proof.

LEMMA 4.2. *There exists a linear space isomorphism* Φ *of* $\mathfrak{X}$ *onto* $\mathfrak{Y}$ *which preserves orthogonality and is such that* $(\mathfrak{D}u)\Phi = (\mathfrak{D}u)F$, *where* F *is the mapping given in Theorem* 2.6 *or* 3.5.

Proof. The existence of a linear space isomorphism $x \rightarrow x\Phi$ such that $(\mathfrak{D}u)\Phi = (\mathfrak{D}u)F$ is given, in view of property (iii) in Theorems 2.6 and 3.5, by a known result from projective geometry.[11] That Φ preserves orthogonality follows from property (ii) of Theorems 2.6 and 3.5.

In the next theorem, which is the desired representation theorem, Φ will denote the linear space isomorphism of $\mathfrak{X}$ onto $\mathfrak{Y}$ given by Lemma 4.2. Also $\mathcal{E}_0$ will denote the group, under the circle operation, consisting of all elements μ in the center of $\mathcal{E}$ such that $\mu \circ \mu^* = 0$.

THEOREM 4.3. *The group isomorphism* $G \rightarrow \mathfrak{g}(G)$ *has the form*

---

[11] For an outline of a proof of this result, see, e.g., [7, Lemma 4.4].

$$(4.1) \qquad \mathfrak{g}(G) = \gamma(G) I \circ (\Phi^{-1} G \Phi),$$

where $G \to \gamma(G)$ *is a homomorphism of* $\mathfrak{G}$ *into* $\mathcal{E}_0$.

Proof. Let T be an arbitrary minimal involution in $\mathfrak{G}$. Then $\Phi^{-1}T\Phi$ is a minimal involution in $\mathfrak{U}(\mathfrak{Y})$ but not necessarily in $\mathfrak{H}$. However the minimal non-isotropic subspace of $\mathfrak{Y}$ associated with $\Phi^{-1}T\Phi$ is the same as that associated with $\mathfrak{g}(T)$. Therefore $\mathfrak{g}(T) = \rho I \circ (\Phi^{-1}T\Phi)$, where $\rho = 0$ or 2. Hence $\mathfrak{g}$ has the form (4.1) for minimal involutions. Now define the mapping

$$(4.2) \qquad G \to G^{\sigma} = \Phi \mathfrak{g}(G) \Phi^{-1},$$

which is clearly an isomorphism of $\mathfrak{G}$ onto another subgroup of $\mathfrak{U}(\mathfrak{X})$. Observe that, if T is a minimal involution in $\mathfrak{G}$, then $T^{\sigma} = \rho I \circ T$, where $\rho = 0$ or 2.

Consider an arbitrary $G \in \mathfrak{G}$ and an arbitrary minimal involution $T \in \mathfrak{G}$. Then $G \circ T \circ G^{\circ} = (I - G)T(I - G)^{-1}$ is also a minimal involution. Therefore $(G \circ T \circ G^{\circ})^{\sigma} = \rho_1 I \circ (G \circ T \circ G^{\circ})$, $\rho_1 = 0$ or 2. But also $(G \circ T \circ G^{\circ})^{\sigma} = G^{\sigma} \circ T^{\sigma} \circ (G^{\sigma})^{\circ}$ and $T^{\sigma} = \rho_2 I \circ T$, $\rho_2 = 0$ or 2. Hence $G^{\sigma} \circ T \circ (G^{\sigma})^{\circ} = \rho I \circ (G \circ T \circ G^{\circ})$, $\rho = \rho_1 \circ \rho_2 = 0$ or 2. This last result can be written in the form

$$\rho I = (I - G^{\sigma})T(I - G^{\sigma})^{-1} + (I - G)T(I - G)^{-1}(\rho - 1).$$

Each term on the right is at most two-dimensional and consequently the right hand side is at most four-dimensional. Since $\dim \mathfrak{X} \geq 6$, it follows that $\rho = 0$. In other words,

$$(4.3) \qquad (I - G^{\sigma})T(I - G^{\sigma})^{-1} = (I - G)T(I - G)^{-1}.$$

Suppose now that there exists a $z \in \mathfrak{X}$ such that $v = z(I - G^{\sigma})$ and $u = z(I - G)$ are linearly independent. We prove that this leads to a contradiction. In the unitary case choose, by Lemma 2.2, a non-isotropic vector w such that $(u, w) = 1$ and $(v, w) = 0$. Define $xT = 2(x, w)(w, w)^{-1}w$. Then $uT \neq 0$ and $vT = 0$. With this choice of T, apply (4.3) to z to obtain $uT(I - G)^{-1} = 0$. Since this implies $uT = 0$, we have a contradiction in the unitary case. In the symplectic case, choose w linearly independent of u, v such that $(u, w) = (v, w) = 0$ and then choose w' such that $(u, w') = (w, w') = 1$ and $(v, w') = 0$. Define $xT = 2(x, w')w - 2(x, w)w'$. Then $uT = 2w$ and $vT = 0$. With this choice of T, apply (4.3) to z to obtain $2w(I - G)^{-1} = 0$. This implies $w = 0$, contrary to the choice of w. We have thus proved that $x(I - G^\sigma)$ and $x(I - G)$ are linearly dependent for all x. A similar proof shows that $x(I - G^\sigma)$ and $x(I - G)$ also vanish simultaneously. Therefore, for every $x \varepsilon \mathfrak{X}$, there exists $\lambda_x \varepsilon \mathfrak{D}$ such that $x(I - G^\sigma) = \lambda_x x(I - G)$ or

$$(4.4) \qquad x(I - G^\sigma)(I - G)^{-1} = \lambda_x x.$$

Since the left hand side of (4. 4) is additive in x, we obtain for all x, y,

$$(4.5) \qquad (\lambda_{x+y} - \lambda_x)x + (\lambda_{x+y} - \lambda_y)y = 0.$$

Since $\dim \mathfrak{X} > 1$, (4. 5) implies that λ_x is equal to a constant λ independent of x. Hence $I - G^\sigma = \lambda(I - G)$ and this can be written in the form $G^\sigma = (1 - \lambda)I \circ G$. Returning to $\mathfrak{g}(G)$, we have

$$(4.6) \qquad \mathfrak{g}(G) = \gamma(G)I \circ (\Phi^{-1}G\Phi),$$

where $\gamma(G) = (1 - \lambda)^\phi$. It remains to prove that $G \rightarrow \gamma(G)$ is a homomorphism of $\mathfrak{G}$ into $\mathcal{E}_0$. It is obvious from the linearity of $\mathfrak{g}(G)$ that $\gamma(G)$ is in the center of $\mathcal{E}$. Moreover, for all $x, y \varepsilon \mathfrak{Y}$,

$$(4.7) \qquad (x\mathfrak{g}(G), y\mathfrak{g}(G)) = (x\mathfrak{g}(G), y) + (x, y\mathfrak{g}(G)).$$

Substituting (4. 6) in (4. 7) with $\gamma = \gamma(G)$, we obtain

$$\begin{aligned}(4.8) \quad \gamma(x, y)\gamma^* &+ \gamma(x, y\Phi^{-1}G\Phi)(1 - \gamma^*) + (1 - \gamma)(x\Phi^{-1}G\Phi, y)\gamma^* \\ &+ (1 - \gamma)(x\Phi^{-1}G\Phi, y\Phi^{-1}G\Phi)(1 - \gamma^*) \\ &= \gamma(x, y) + (1 - \gamma)(x\Phi^{-1}G\Phi, y) + (x, y)\gamma^* + (x, y\Phi^{-1}G\Phi)(1 - \gamma^*).\end{aligned}$$

Using the fact that $\Phi^{-1}G\Phi \varepsilon \mathcal{U}(\mathfrak{Y})$ and that γ is in the center of $\mathcal{E}$ and then collecting terms in (4. 8), we finally obtain $\gamma \circ \gamma^* = 0$. Hence $\gamma \varepsilon \mathcal{E}_0$. It is obvious that $G \rightarrow \gamma(G)$ is a homomorphism of $\mathfrak{G}$ into $\mathcal{E}_0$; therefore the proof is complete.

COROLLARY. *If $\mathfrak{G}$ coincides with its commutator subgroup, then $\gamma(G) \equiv 0$. In particular, if $\mathfrak{X}$ is symplectic and finite-dimensional, then $\gamma(G) \equiv 0$* [4, p. 12].

Proof. Since $\mathcal{E}_0$ is commutative, γ maps commutators of $\mathfrak{G}$ into 0.

COROLLARY. *If $\mathcal{E}$ is a field with the identity mapping as involution, then $\gamma(G) = 0$ or 2 for all $\mathfrak{G}$.*

Proof. In this case we have $\gamma(G) \circ \gamma(G) = 0$ which implies $\gamma(G) = 0$ or 2.

We close with a remark on the unitary case when the vector spaces have index zero; that is, every vector is non-isotropic. This case is of course included in the above treatment; however the discussion becomes considerably simpler. Moreover it is possible to weaken the dimension condition from six to three and drop the condition which excludes the field F_3.

YALE UNIVERSITY.

BIBLIOGRAPHY.

[1] J. Dieudonné, "On the automorphisms of the classical groups," *Memoirs of the American Mathematical Society*, No. 2 (1950).

[2] ———, "Sur les automorphismes des groupes classiques," *Comptes Rendus de l'Académie des Sciences (Paris)*, vol. 225 (1947), pp. 914-915.

[3] ———, "Sur les automorphismes du group unitaire," *ibid.*, vol. 225 (1947), pp. 975-976.

[4] ———, "Sur les groupes classiques," *Actualités Scientifiques et Industrielles*, No. 1040 (1948).

[5] L. K. Hua, "On the automorphisms of the symplectic group over any field," *Annals of Mathematics*, vol. 49 (1948), pp. 739-759.

[6] N. Jacobson, "On the theory of primitive rings," *ibid.*, vol. 48 (1947), pp. 8-21.

[7] C. E. Rickart, "Isomorphic groups of linear transformations," *American Journal of Mathematics*, vol. 72 (1950), pp. 451-464.

ON THE IRREDUCIBILITY OF POLYNOMIALS WITH LARGE THIRD COEFFICIENT. II.*

By ALFRED BRAUER.

This paper is a continuation of my paper "On the irreducibility of polynomials with large third coefficient," this JOURNAL, vol. 70 (1948), pp. 423-432. The enumeration of the theorems and equations will be continued here. Combining the main results of the first paper and new results we will prove the following theorem.

THEOREM 5. *Let* $f(x) = x^n + a_1x^{n-1} + \cdots + a_n$ $(a_n \neq 0)$ *be a polynomial with integral rational coefficients. Let* m *and* m^* *be the minimum of the partial sums of the series* $0 + a_3 + a_4 + \cdots + a_n$ *and* $0 - a_3 + a_4 - + \cdots + (-1)^n a_n$, *respectively.*

Assume that one of the following conditions is satisfied:

(α) $a_2 \geqq a_1^2/4 + |m^*|$ for $n > 2$ and $a_1 > 0$,

(β) $a_2 \geqq a_1^2/4 + |m|$ for $n > 2$ and $a_1 < 0$,

(γ) $|a_1| \leqq 3$,

(δ) The inequality

$$|a_\nu| \geqq \tfrac{1}{4}\binom{n-2}{\nu-2}\{|a_1| + n - 2\}^2 + \binom{n-2}{\nu-1}\{|a_1| + n - 2\} + \binom{n-2}{\nu}$$

holds for at least one value of ν $(\nu = 2, 3, \cdots, n)$.

We set $t = |1 + a_4| + |a_1 + a_3| + |a_5| + |a_6| + \cdots + |a_n|$. If

$$a_2 > t, \tag{34}$$

then $f(x)$ is irreducible in the field of rational numbers.

Proof. The cases (α) and (β) follow from the proof of Theorem 1a. It follows from (34) as in the proof of Theorem 1 that $f(x)$ has $n - 2$ roots in the interior and 2 roots in the exterior of the unit circle, and for the irreducibility of $f(x)$ it is sufficient to prove that the two roots in the exterior of the unit circle are imaginary roots.

---

* Received July 16, 1950.

Let us assume that these two roots are real. We denote them by x_1 and x_2, and the other roots by $x_3, x_4, \cdots, x_n$.

Assume now that (γ) holds. For $a_1 = 0$ the statement follows from Theorem 2, but it will be obtained here again. It is sufficient to prove that $f(x)$ has no real root greater than or equal to 1. It follows then that $f(-x)$ has no such root either since its coefficients satisfy (γ) and (34). Therefore $f(x)$ has no real root less than or equal to -1, and x_1 and x_2 are imaginary roots.

We have to consider some cases.

1). $a_1 \geqq 0$. It follows from (34) that

$$a_2 > | a_4 | - 1 + | a_3 | - | a_1 | + | a_5 | + | a_6 | + \cdots + | a_n |,$$

hence for $x \geqq 1$

$$a_2 x^{n-2} > -x^{n-2} - a_1 x^{n-2} + (| a_3 | + | a_4 | + \cdots + | a_n |) x^{n-2},$$

$$x^n + a_1 x^{n-1} + a_2 x^{n-2} > (| a_3 | + | a_4 | + \cdots + | a_n |) x^{n-2},$$

$$f(x) \geqq x^n + a_1 x^{n-1} + a_2 x^{n-2} - (| a_3 | + | a_4 | + \cdots + | a_n |) x^{n-2} > 0.$$

2). $-1 \geqq a_1 \geqq -3$ and $a_3 \leqq 0$. It follows from (34) that

$$a_2 - 1 \geqq | a_1 | + | a_3 | + | a_4 | - 1 + | a_5 | + \cdots + | a_n |,$$

hence for $x \geqq 1$

$$(35) \qquad a_2 x^{n-2} \geqq (| a_1 | + | a_3 | + | a_4 | + \cdots + | a_n |) x^{n-2}.$$

On the other hand

$$(36) \qquad x^n - | a_1 | x^{n-1} + | a_1 | x^{n-2} = x^{n-2}(x^2 - | a_1 | x + | a_1 |) > 0,$$

since $a_1 = -1, -2$ or -3. By adding (35) and (36) we obtain

$$x^n + a_1 x^{n-1} + a_2 x^{n-2} > (| a_3 | + | a_4 | + \cdots | a_n |) x^{n-2},$$

hence $f(x) > 0$ for $x \geqq 1$.

3). $-1 \geqq a_1 \geqq -3$ and $a_3 > 0$.

We have for $x \geqq 1$

$$(37) \qquad x^n - 3x^{n-1} + x^{n-2} + 3x^{n-3} - x^{n-4} = x^{n-4}\{x^2(x-2)^2 + (x-1)^3\} > 0$$

and

$$(38) \qquad x^{n-1} - x^{n-3} \geqq 0.$$

It follows from (37) and (38) that

$$(39) \qquad x^n + a_1x^{n-1} + x^{n-2} + |a_1|x^{n-3} - x^{n-4} > 0$$

for $-1 \geqq a_1 \geqq -3$. Moreover by (34)

$$(40) \quad a_2 - 1 \geqq |a_1| - |a_3| + |1 + a_4| + |a_5| + |a_6| + \cdots + |a_n|,$$

$$(41) \quad (a_2-1)x^{n-2} \geqq (a_2-1)x^{n-3} \geqq (|a_1| - |a_3|)x^{n-3} + |1+a_4|x^{n-4} + (|a_5| + |a_6| + \cdots + |a_n|)x^{n-4}.$$

Adding (41) and (39) we obtain

$$x^n + a_1x^{n-1} + a_2x^{n-2} + a_3x^{n-3} > (|a_4| + |a_5| + \cdots + |a_n|)x^{n-4}.$$

Hence $f(x) > 0$ for $x \geqq 1$, and the theorem is proved if (γ) holds.

If $a_1 = 4$, then the polynomial may have real roots in the exterior of the unit circle. For instance, $x^5 - 4x^4 + 2x^3 + 4x^2 - x - 1$ is negative for $x = 2$, although its coefficients satisfy (34).

Now we assume that (δ) holds. It follows from (34) that

$$\begin{aligned} f(1) &= 1 + a_1 + a_2 + \cdots + a_n \\ &\geqq a_2 - |1 + a_4| - |a_1 + a_3| - |a_5| - |a_6| - \cdots - |a_n| \\ &= a_2 - t > 0. \end{aligned}$$

Therefore either no root or two roots are greater than 1, hence sign x_1 = sign x_2, hence

$$(42) \qquad |x_1x_2| \leqq |x_1 + x_2|^2/4.$$

Let us denote the κ-th elementary symmetric function of $x_3, x_4, \cdots, x_n$ by C_κ. Since $|x_\rho| < 1$ for $\rho = 3, 4, \cdots, n$, we have

$$(43) \qquad |C_\kappa| < \binom{n-2}{\kappa} \qquad (\kappa = 1, 2, \cdots, n-2).$$

On the other hand $x_1 + x_2 = -a_1 - C_1$, hence, by (43) and 42), respectively,

$$(44) \quad |x_1 + x_2| < |a_1| + n - 2, \qquad (45) \quad |x_1x_2| < (|a_1| + n - 2)^2/4.$$

Moreover $(-1)^\nu a_\nu = x_1x_2C_{\nu-2} + (x_1 + x_2)C_{\nu-1} + C_\nu$ for $\nu = 2, 3, \cdots, n$ and by (43), (44), and (45)

$$|a_\nu| < \tfrac{1}{4}\binom{n-2}{\nu-2}\{|a_1| + n - 2\}^2 + \binom{n-2}{\nu-1}\{|a_1| + n - 2\} + \binom{n-2}{\nu}.$$

Since for a certain ν the inequality (δ) holds, a contradiction is obtained.

Therefore x_1 and x_2 are imaginary roots. This completes the proof of the theorem.

Our result can be formulated also as follows:

THEOREM 5a. *If a polynomial with integral rational coefficients satisfies the assumptions of* Theorem 5, *then it has a pair of complex roots in the exterior of the unit circle while all the other roots lie in the interior of the unit circle.*

A similar theorem can be obtained for polynomials with arbitrary real coefficients.

It was shown in the first paper that (34) gives the best possible bound for a_2 if $f(x) = x^5 - 4x^4 + Kx^3 + 4x^2 - 2x - 2$. For these polynomials the condition (β) holds if $K > t$.

Moreover, it was stated there that (β) holds for the polynomials

$$f(x) = x^n - a_1x^{n-1} + a_2x^{n-2} - a_3x^{n-3} - \cdots - a_n \tag{46}$$

where all the a_ν are non-negative, a_4 and a_n positive and

$$| a_1 | \leqq 2t^{1/2}. \tag{47}$$

But this is not correct since $m \neq 0$. However, it follows now that (34) gives the best bound for a_2 for the polynomials (46) if instead of (47) either (γ) or (δ) holds. In these cases the polynomials are irreducible for $a_2 > t$ while they are reducible for $a_2 = t$ since $f(1) = a_2 - t = 0$.

In the formulation of Theorem III in the first paper a factor s^2 is erroneously omitted. It should read "or if $a_2 \geqq (7/2)^{2n-2}s^2$."

UNIVERSITY OF NORTH CAROLINA.

SUPPLEMENT TO A PAPER OF G. DE B. ROBINSON.*

By R. M. Thrall and G. de B. Robinson.

This note is a supplement to a paper of one of the authors.[1] Its purpose is to correct a formula in that paper, and to derive a new formula which depends on this corrected form.

1. The revisions. The portion of SG_3, § 7, which appears on p. 291 should read as follows:

But we can go deeper still. Let $S_b{}^*$ be the symmetric group of degree b and let W^* be an element of $S_b{}^*$ with cycle structure b_1 1-cycles, b_2 2-cycles, $\cdots$, b_k k-cycles. Then let W^{*+} be an element of S_n $(n = a + bq)$ with a 1-cycles, b_1 q-cycles, b_2 $2q$-cycles, $\cdots$. b_k kq-cycles, and let V be any permutation on the fixed letters of W^{*+}. We prove the following generalization of Theorem 6. 5.

$$\chi_\alpha(VW^{*+}) = \sigma_\alpha \cdot (\chi(W^*) \text{ in } [\alpha]_q{}^*) \cdot \chi_{\alpha^0}(V). \tag{7. 1r}$$

Now we can write

$$\begin{aligned}\chi_\alpha(VW^{*+}) &= \Sigma\chi_\gamma(V) \cdot \chi_\alpha{}^\gamma(P_b \cdot W) \\ &= \Sigma\chi_\gamma(V) \cdot \chi_{(b_1q)}\chi_{(b_2q)} \cdots \chi_{(b_kq)}.\end{aligned} \tag{7. 2r}$$

As before, we are only interested in those terms which arise from a sequence of skew hook representations (b_iq), so that we can take $[\gamma] = [\alpha^0]$. Consider one such term on the right hand side of 7. 2r and apply 5. 4 to its component hooks. Here $\sigma = \sigma_\alpha$ is the product of the parities of the hooks of length q, σ' is the product of the parities of the b_iq-hooks, and σ^* is the product of the parities of the corresponding b-hooks in $[\alpha]_q{}^*$. I. e. multiplying the k equations 5. 4 we obtain

$$\sigma' = \sigma\sigma^*. \tag{7. 3}$$

It follows from 5. 2 that σ is the same for all such terms of 7. 2r, so we can sum 7. 3, applying the Murnaghan-Nakayama recursion formula iterated k-times to yield the equation

---

* Received May 1, 1950.

[1] G. de B. Robinson, "On the representations of the symmetric group" (third paper), *American Journal of Mathematics*, vol. 70 (1948), pp. 277-294. We shall refer to this paper as SG_3.

7.4r $$\chi_{\alpha}^{\alpha^0}(W^{*+}) = \sigma_{\alpha} \cdot (\chi(W^*) \text{ in } [\alpha]_q^*),$$

which proves 7.1r.

As before, if $[\alpha']$ is a second diagram for n with q-core containing $a' > a$ nodes then we conclude that

7.5r $$\chi_{\alpha'}(VW^{*+}) = 0.$$

These changes also apply to the paragraph of SG_3 beginning on the bottom of p. 278 and in particular to formula (7) p. 279.

2. The new formula. We are interested in the frequency τ_{α} of the identity representation of S_b^* in $[\alpha]_q^*$. On the one hand this frequency (SG_3, p. 284, last paragraph) is one or zero according as each disjoint constituent of $[\alpha]_q^*$ does or does not consist of a single row. On the other hand we can get this frequency from the orthogonality relations for characters of S_b^*. Thus we have

$$(\Sigma\chi(W^*) \text{ in } [\alpha]_q^*)/b\,! = \tau_{\alpha},$$

where the sum is over all W^* in S_b^*. Now sum 7.2r over W^*, substitute the above formula and we get

(7.9) $$\Sigma\chi_{\alpha}(VW^{*+})/b\,! = \tau_{\alpha}\sigma_{\alpha}\chi_{\alpha^0}(V).$$

Similarly, if $[\alpha']$ has q-core with more than a nodes we have on summing (7.5r) over W^* that

(7.10) $$\Sigma\chi_{\alpha'}(VW^{*+}) = 0.$$

The case $n = bq$ is of especial interest. Then we shall prove that for every diagram $[\alpha]$ with n nodes

(7.11) $$\sum_{W^* \varepsilon S_b^*} \chi_{\alpha}(W^{*+})/b\,! = \theta_{\alpha},$$

where θ_{α} is zero if the q-core of α is not zero or if $[\alpha]$ has more than q-rows, and $\theta_{\alpha} = \sigma_{\alpha}$ if the q-core of $[\alpha]$ is zero and $[\alpha]$ has q or less rows.

Proof. When $n = bq$ we have $a = 0$ and hence (7.10) applies for all $[\alpha]$ whose q-core is not zero. For the $[\alpha]$ with q-core zero we apply (7.9) and the formula (7.11) will be established as soon as we can show that for $[\alpha]$ with zero q-core τ_{α} is 0 or 1 according as $[\alpha]$ has $k > q$ or $k \leqq q$ rows.

We first state a criterion for a diagram of $n = bq$ nodes to have zero core. Let $[\alpha]$ correspond to the partition $\alpha_1 + \cdots + \alpha_k + \alpha_{k+1} + \cdots + \alpha_m = n$, where $\alpha_1 \geqq \alpha_2 \geqq \cdots \geqq \alpha_k > 0 = \alpha_{k+1} = \cdots = \alpha_m$ and $m \geqq k$, and let $l_j = \alpha_j + m - j$, $j = 1, \cdots, m$.

7.12. *Let m be a multiple of q. Then the diagram $[\alpha]$ has zero q-core if and only if exactly m/q of the l_j fall into each residue class modulo q.*

This criterion is implicit in the Murnaghan-Nakayama recursion formula [2] as well as in SG_3 and papers of Todd [3] and Staal.[4] We give here a direct proof.

Suppose that $[\alpha']$ is the diagram obtained by removing from $[\alpha]$ the rim of a q-hook which begins in the i-th row and ends in the j-th row. Then for

$$\begin{array}{lll} h < i & \alpha'_h = \alpha_h & l'_h = l_h \\ i \leq h < j & \alpha'_h = \alpha_{h+1} - 1 & l'_h = l_{h+1} \\ h = j & \alpha'_j = \alpha_i - q + j - i & l'_j = l_i - q \\ h > j & \alpha'_h = \alpha_h & l'_h = l_h. \end{array}$$

Thus we see that for any fixed m the congruence class mod q for the l_j are the same as for the l'_j. If s removals of q-hooks from $[\alpha]$ give $[\alpha^{(s)}]$ we see that the congruence classes mod q for the l_j are the same as for the $l_j^{(s)}$. If, in particular, $n = sq$ then $[\alpha^{(s)}]$ is the zero diagram, $l_j^{(s)} = m - j$, and hence the $l_j^{(s)}$ fall into q congruences classes each containing m/q elements. This establishes the necessity of the criterion.

For any m there will be a q-hook beginning at the end of the i-th row if and only if $l_i \geq q$ and $l_i - q$ is not equal to any l_h for $h > i$. For, if $l_i - q$ is non negative and different from the other l_h we can find j such that $l_j > l_i - q > l_{j+1}$. The corresponding α_j must be positive (i. e. $j \leqq k$) since $2 \leq l_j - l_{j+1} = \alpha_j - \alpha_{j+1} + 1$ or $\alpha_j \geqq \alpha_{j+1} + 1$. Moreover, the q-th node down along the rim starting from the end of the i-th row will lie in the j-th row and beyond the node above the end of the $(j+1)$-th row, which shows that a q-hook can be removed. If $l_i - q$ is negative or equal to some l_j then either $j \leq k$ and the q-th node down from the end of the i-th row falls in the $(j-1)$-th row directly above the last node of the j-th row or $j > k$ and there are less than q-nodes available from which to form a hook.

Under the hypothesis of the criterion we have $l_{k-1} = m - k - 1$ and $l_k = \alpha_k + m - k$. There must be at least one $h < k$ for which $l_h \equiv m - k$. Let i be the largest such index. Then $l_i = \alpha_i + (k - i) + (m - k)$ so that $l_i - q \geq m - k$, i. e. $l_i - q \neq l_h$ for $h > k$. Because of the maximal nature of i we cannot have $l_i - q = l_h$ for $i < h \leqq k$. Hence, there is a q-hook beginning at the end of the i-th row.

---

[2] F. D. Murnaghan, "On the representations of the symmetric group," *American Journal of Mathematics*, vol. 59 (1937), pp. 437-488.

T. Nakayama, "On some modular properties of irreducible representations of the symmetric group, Part I," *Japanese Journal of Mathematics*, vol. 17 (1940), pp. 165-184.

[3] J. A. Todd, "A note on the algebra of S-functions," *Proceedings of the Cambridge Philosophical Society*, vol. 45 (1949), pp. 328-334.

[4] R. A. Staal, "Star diagrams and the symmetric group," *Canadian Journal of Mathematics*, vol. 2 (1950), pp. 79-92.

Now, that we have shown that the congruence conditions of the criterion guarantee the existence of at least one q-hook it follows from the congruence of the l_j and the l_j' that $[\alpha']$ will also have at least one hook and thus finally leads to the conclusion that $[\alpha]$ has zero q-core, i. e. shows that the criterion affords a sufficient condition for a zero core.

We now return to the determination of τ_α. Let $[\alpha]$ have q-core zero and choose s so that $m = sq \geq k > (s-1)q$. Then none of the non-negative integers l_h for $h > k$ will be as large as $q-1$, hence exactly s of the l_h with $h \leq k$ will be congruent to $q - 1 \bmod q$. Let these be $l_{i_1} < l_{i_2} < \cdots < l_{i_s}$. We have seen above that there is a q-hook Q_1 beginning at the end of the i-th row and since the l_{i_j} are all congruent mod q, there will be for each j a $b_j q$-hook Q_j beginning at the end of the i_j-th row and ending with the last node of Q_1.

Now it follows from[5] Theorem 4.7 of SG_3 that whenever $s > 1$ some disjoint constituent of $[\alpha]_q{}^*$ will have more than one row. On the other hand if $k \leq q$ (i. e., if $s = 1$) no two l_h for $h \leq k$ can be congruent and hence no disjoint constituent of $[\alpha]_q{}^*$ can contain more than one row. This completes the proof of Theorem 7.11.

The arguments used in the proof of Theorem 7.12 serve also to give a short proof of the uniqueness of the q-core (first proved by Nakayama[6]). Rephrasing the criterion that a diagram have a q-hook we get the following theorem:

7.13. *A diagram is a q-core if and only if each class of congruent l_j's (formed for any m) contains all smaller non-negative integers congruent to the largest one in the class.*

Consider two q-cores $[\alpha]$ and $[\alpha']$ and select any $m \geqq \max(k, k')$. Then if each congruence class mod q for the l_j has the same order as the corresponding class for the l_j', it follows from 7.13 that $[\alpha] = [\alpha']$. But since the congruence classes mod q are the same for a diagram $[\beta]$ as for any diagram obtained from $[\beta]$ by removing q-hooks we conclude that $[\beta]$ can have only one q-core.

UNIVERSITY OF MICHIGAN
AND
UNIVERSITY OF TORONTO.

---

[5] See also R. A. Staal, *loc. cit.*, Theorem C.

[6] T. Nakayama, *loc. cit.*

CENTRAL LIBRARY UNIVERSITY

A REPRESENTATION FOR BOOLEAN ALGEBRAS.*

By Franklin Haimo.

1. The content. The success of Stone [5] in representing Boolean algebras by fields of sets leads one to search for other representations. (See also [2, p. 159], [4].) Using splitting endomorphisms on a type of Abelian group with an order relation, we are led to a faithful representation of a given Boolean algebra $\mathfrak{B}$. The group employed will be called a Boolean group. A special case of a Boolean group is a Boolean ring, and we shall show that $\mathfrak{B}$ can be represented by a set of endomorphisms on its corresponding Boolean ring. If $\mathfrak{B}$ is complete the group on which it operates will be lower complete and upper conditionally complete. The Boolean groups are themselves special cases of a wider class of groups with an order relation, the vector ordered groups. Vector ordered groups generalize direct sums of Abelian groups, and their splitting endomorphisms form Boolean algebras.

A simple example illustrates virtually the entire theory: Let $\mathfrak{J}_1$ be the ring of integers modulo 2, and let $\mathfrak{G}$ be the Cartesian product of some infinite collection of copies of $\mathfrak{J}_2$. $\mathfrak{G}$ can be made into the strong direct sum of the $\mathfrak{J}_2$ by introducing component-wise addition and multiplication. Then $\mathfrak{G}$ becomes a Boolean ring with a unity. It is also a Boolean algebra [2, p. 154] if one writes $x \geq y$ whenever each component of $y \in \mathfrak{G}$ equals the corresponding component of $x \in \mathfrak{G}$ or is zero. Let $\mathfrak{P}(\mathfrak{G})$ be the set of all endomorphisms V of the group structure of $\mathfrak{G}$ for which $V^2 = V$ and $V(x) \leq x$ for every $x \in \mathfrak{G}$. If an order relation is defined in $\mathfrak{P}$ in the obvious way, $\mathfrak{P}$ becomes a Boolean algebra isomorphic to the Boolean algebra $\mathfrak{G}$, so that the algebra $\mathfrak{G}$ is represented by a set of operators on the group $\mathfrak{G}$. If we replace each summand $\mathfrak{J}_2$ by $\mathfrak{J}$, the group of integers, the corresponding strong direct sum group $\mathfrak{H}$ can be ordered component-wise as was done for $\mathfrak{G}$. Then $\mathfrak{P}(\mathfrak{H}) \cong \mathfrak{P}(\mathfrak{G})$. In fact, we can reconstruct both $\mathfrak{H}$ and $\mathfrak{G}$ from $\mathfrak{P}$ by assigning coefficients from $\mathfrak{J}$ (respectively, from $\mathfrak{J}_2$) to the atoms of $\mathfrak{P}$. The groups $\mathfrak{G}$ and $\mathfrak{H}$ are examples of Boolean groups, to be defined below. If the members of a proper subcollection of the summands $\mathfrak{J}_2$ are replaced by $\mathfrak{J}$'s and if the remaining summands are unaltered, the resulting strong direct sum group $\mathfrak{K}$, ordered as above, has the property $\mathfrak{P}(\mathfrak{K}) \cong \mathfrak{P}(\mathfrak{G})$. $\mathfrak{K}$ is an example of a vector ordered group of a type which we shall designate as "strong." The

---

* Received November 15, 1949; revised March 9, 1951.

use of the atoms of $\mathcal{P}$ is inessential, as we shall see below. For instance, had we reduced the group $\mathfrak{G}$ modulo the subgroup $\mathfrak{G}^w$, the weak direct sum of the same summands, we would have been led to an atomless example of the same theory. Part I of this paper will be concerned with the (strong) vector ordered groups. In Part II we shall prove the representation theorems, using results from Part I when needed.

2. **Notation.** Groups, subgroups and "complexes" will be denoted by Gothic capitals; their elements, however, will be given by small Latin letters, their endomorphisms by Latin capitals. Although early Greek small letters are used in Part I for indexing, in Part II they usually denote elements of Boolean algebras. These latter algebras and some other partly ordered sets will often be given by Latin script capitals. Exceptions to the above conventions and additional notation will be clear from the context. The basic terminology is essentially that of G. Birkhoff [2].

Part I.

3. **The concepts $\perp$ and $||$.** Let $\mathcal{S}$ be a partly ordered set with universal infimum 0. If $x, y \varepsilon \mathcal{S}$, we say that $x \perp y$ (x is perpendicular to y) if (1) $z \varepsilon \mathcal{S}$, $z \leqq x, z \leqq y$ imply $z = 0$, and (2) there exists $w \varepsilon \mathcal{S}$ such that $x \leqq w, y \leqq w$. It is clear that $\perp$ is a symmetric relation on $\mathcal{S}$ and that $0 \perp x$ for every $x \varepsilon \mathcal{S}$. $x \perp x$ if, and only if, $x = 0$. For subsets $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{S}$, we say that $\mathcal{A} \perp \mathcal{B}$ if $a \varepsilon \mathcal{A}, b \varepsilon \mathcal{B}$ imply $a \perp b$.

We say that $x \parallel y$ (x is parallel to y) if (1) $z \leqq x$, $z \perp y$ imply $z = 0$, and (2) $w \perp x$, $w \leqq y$ imply $w = 0$. $\parallel$ is a reflexive and symmetric relation on $\mathcal{S}$. $\mathcal{A} \parallel \mathcal{B}$ is defined in the obvious way.

4. **Vector ordered groups.** Let $\mathfrak{G}$ be an additive Abelian group, partly ordered by a relation $\leqq$, (where $x \leqq y$ may be written $y \geqq x$), subject to the following axioms:

(1) $x \geqq 0$, for every $x \varepsilon \mathfrak{G}$ (where 0 is the additive identity of $\mathfrak{G}$).

(2) $x \geqq y$ implies $y \perp (x - y)$.

(3) $x \perp y$ implies $x + y \geqq x, y$.

(4) $x + y \geqq x, y$ imply that $x + y = x \cup y$, the least upper bound (l. u. b.) of x and y.

(5) For fixed $a \varepsilon \mathfrak{G}$ the sets $\mathfrak{K}_a = [x \mid x \varepsilon \mathfrak{G}; x \perp a]$ and $\mathfrak{H}_a = [x \mid x \varepsilon \mathfrak{G}; z \leqq x, z \perp a \text{ imply } z = 0]$ are subgroups of $\mathfrak{G}$.

(6) (a) To each $x \varepsilon \mathfrak{G}$ there exists at least one maximal element m of $\mathfrak{G}$ with $x \leqq m$. (b) Given $a, b \varepsilon \mathfrak{G}$, there exists $c \varepsilon \mathfrak{G}$ such that $c \leqq a$,

$c \perp b$, and such that $d \leq a$, $d \perp b$ imply $d \leq c$. We write $c = a_{\perp b}$. (c) Given $a, b \varepsilon \mathfrak{G}$. there exists $c \varepsilon \mathfrak{G}$ such that $c \leq a$, $c \leq b$, and such that $d \leq a$, $d \leq b$ imply $d \leq c$. We write $c = ab$.

Abelian groups satisfying (1) through (6) are called vector ordered groups (v. o. groups). For instance, let $\mathfrak{G}$ be the direct sum of a finite number of additive Abelian groups $\mathfrak{G}_i$, $i = 1, 2, \cdots, n$. If $a, b \varepsilon \mathfrak{G}$, we say that $a \geq b$ if b can be obtained from a by replacing some (or no) non-zero components of a by zeros, leaving the other components fixed. It is clear that this ordering make $\mathfrak{G}$ a v. o. group.

Axiom (6) (c) will seldom be used and is given mainly for completeness. We shall often replace (6) by:

(6′) Every chain in $\mathfrak{G}$ has a l. u. b. in $\mathfrak{G}$.

Abelian groups satisfying (1) through (5) and (6′) are called strong v. o. groups. We shall show that (6′) and the other axioms imply (6) so that a strong v. o. group is a v. o. group. If one orders a strong direct sum $\mathfrak{G}$ of a set of additive Abelian groups $\mathfrak{G}_a$ in the same fashion as direct sums were ordered above, then $\mathfrak{G}$ becomes a strong v. o. group. Proofs will be carried through for v. o. groups rather than for strong v. o. groups except where the contrary is indicated.

5. Elementary results. The following have completely trivial proofs. Often these results will be used without explicit reference. Those, the proofs of which employ (6), are starred. (a) $x + y \geq x$ is equivalent to $x + y \geq y$. (b) A non-zero, non-maximal element of $\mathfrak{G}$ is $\perp$ to some other non-zero, non-maximal element. (c) The following are equivalent: $x \geq y$, $-x \geq -y$ and $x \geq x - y$. (d) $x + y \geq x$ implies $x \perp y$. (e) $x \perp y$, $y \geq z$ imply $x \perp z$. (f) $x + y \geq z$, $x \perp z$ imply $y \geq z$. (g) $u \leq x + y$, $u \leq y$ and $v \leq x$ imply $u \perp v$. (h) $x \perp y, x, y \leq w$ imply $x + y \leq w$. (i) $w \geq x + y$, $x \perp y$ imply $w \geq x, y$. (j) $x \perp y, x \geq z$ imply $x + y \geq z + y$. (k) $x \perp y$, $x \geq z$, $y \geq w$ imply $x + y \geq z + w$. (l) $x \geq y \geq z$ imply $x - z \geq y - z$. (m) The principal of limited homogeneity: if $a \leq b$, then $a + x \leq b + x$ if, and only if, $x \perp (b - a)$. (n) For every integer n, $x \geq y$ implies $nx \geq ny$. (n′) Let $\mathfrak{F}$ be the subgroup of those elements of $\mathfrak{G}$ which are zero or have finite order, the torsion subgroup of $\mathfrak{G}$. Then $\mathfrak{F}$ is convex in $\mathfrak{G}$. That is, $x \varepsilon \mathfrak{G}$, $f \varepsilon \mathfrak{F}$; $x \leq f$ imply $x \varepsilon \mathfrak{F}$. (o) For given $a \varepsilon \mathfrak{G}$, $\mathfrak{K}_a$ and $\mathfrak{H}_a$ are convex, and $\mathfrak{K}_a \cap \mathfrak{H}_a = (0)$, the subgroup of $\mathfrak{G}$ consisting of the element 0 alone. (p) $[x \mid x \varepsilon \mathfrak{G}; x \geq a] = a + \mathfrak{K}_a$, the addition being that of complexes. (q) If $x, y \leq z$, there exist $y', x' \varepsilon \mathfrak{G}$ such that $y' \perp x, x' \perp y$ and $y - x = y' - x'$. (r) For $a \neq 0$, $\mathfrak{K}_a$ contains no maximal element of $\mathfrak{G}$. (s*) If a subset of $\mathfrak{K}_a$

has a l. u. b. in $\mathfrak{G}$, then this l. u. b. is in $\mathfrak{K}_a$. (t) $c \geqq a, b$ and v maximal with respect to the property $v \leqq a, b$ imply $a - v \perp b$ and $b - v \perp a$. (t') If v is maximal with respect to the property $v \leqq a, b$, then $(a - v) \perp b$ if, and only if, there exists $c \in \mathfrak{G}$, $c \geqq a - v$, $c \geqq b$. (u^*) to each non-zero, non-maximal element $x \in \mathfrak{G}$ there exists at least one $x' \in \mathfrak{G}$, non-zero and non-maximal with $x' \perp x$ and x' maximal with respect to this property. (Taking an m from (6)(a), let $x' = m - x$.) (u'^*) In (u^*), $w \perp x'$, $w \neq x$ imply $w \not\geqq x$. (u''^*) If n is an integer, $nx \neq x$ implies $nx \not\geqq x$.

We can define a new binary operation $+'$ on some of the pairs of elements of $\mathfrak{G}$ as follows: if $x \perp y$, let $x +' y = x + y$. $+'$ is an associative operation where it is applicable. Under $+'$, $\mathfrak{G}$ becomes an ordered (cf. (j), (m) above) groupoid, similar, in some respects, to the groupoid of Brandt [3]. The properties of this groupoid will be discussed elsewhere.

We shall use the axiom of choice in the form of Zorn's Lemma [2, p. 42] whenever it is needed.

LEMMA 1. *The first five axioms and* (6′) *imply* (6) *and a stronger form of* (6)(c):

(6)(c′) *Strong vector ordered* $\mathfrak{G}$ *is lower complete.*

Proof. (6)(a) is immediate from Zorn's Lemma. As for (6)(b), suppose that $a, b \in \mathfrak{G}$. By Zorn's Lemma and (6′), there exists $s \in \mathfrak{G}$, $s \leqq a$, $s \perp b$, and s is maximal (see (s^*) above) with respect to this property. If $d \leqq a$, $d \perp b$, Zorn's Lemma supplies $w \leqq d, s$ with w maximal with respect to this property. By (t) above, $s \perp d - w$, and $s + d - w \geqq s$. Since $s, d, w \in \mathfrak{K}_b$, $s - d + w \perp b$. $w \leqq s \leqq a$ imply $a \geqq s - w$. But $a \geqq d$, and, by (t), $d \perp s - w$. Hence $a \geqq s + d - w$. By the maximal character of s, $s = s + d - w$, and $w = d$, $d \leqq s$, proving the maximum property of s.

To prove (6)(c′), let $\mathfrak{C}$ be a non-void class of elements of $\mathfrak{G}$. The class $\mathfrak{L}$ of lower bounds for all the elements of $\mathfrak{C}$ is non-void sinve $0 \in \mathfrak{L}$. For a given chain $\mathfrak{D} \subset \mathfrak{L}$, (6′) provides a l. u. b. which is seen to be in $\mathfrak{L}$. By Zorn's Lemma, to each $j \in \mathfrak{L}$, there exists $c(j) \in \mathfrak{L}$, where $c(j) \geqq j$, and where $c(j)$ is maximal in $\mathfrak{L}$. If c and d are two such maximal elements of $\mathfrak{L}$, then it is easy to show that $c \parallel d$. By an argument similar to the one on $\mathfrak{C}$ we can construct an element v where $v \leqq c, d$ and is maximal with respect to this property. c and d have at least one common upper bound. Applying (t) above, $c - v \perp d$. But $c \geqq v$ implies $c \geqq c - v$. Since $c \parallel d$, $c - v = 0$ and $c = v$. Hence $c \leqq d$. Similarly, $d \leqq c$ and $c = d$. This establishes c as the g. l. b. of the elements of $\mathfrak{C}$, and $\mathfrak{G}$ is thus lower complete.

6. Endomorphisms and projections. Let $\mathfrak{G}$ be a v. o. group. Consider the following five properties of endomorphisms V. (α) $V(x) \leq x$ for every $x \varepsilon \mathfrak{G}$. ($\beta$) $V^2 = V$ (V is splitting on $\mathfrak{G}$). (γ) Let the kernel of V be denoted by $\mathfrak{K}(V)$ and set $\mathfrak{H}(V) = [x \mid x \varepsilon \mathfrak{G};\, V(x) = x]$. Then the subgroups $\mathfrak{H}(V)$ and $\mathfrak{K}(V)$ are elementwise $\perp$. (δ) $x \geq y$ implies $V(x) \geq V(y)$. (ϵ) $x \perp y$ implies $V(x) \perp V(y)$. It is immediate that (δ) and (ϵ) are equivalent, that (α) implies (γ) and (δ) and that (β) and (γ) as a pair imply (α). Likewise, (α) implies (β); for, $V(x) \geq V^2(x)$ so that $V(x) \geq V(x) - V^2(x)$, and $x \geq V(x)$ implies $V(x) \perp x - V(x)$. Therefore, $V(x) - V^2(x) \perp x - V(x)$. But $V(x) - V^2(x) \leq x - V(x)$. This shows that $V(x) - V^2(x) = 0$ for every $x \varepsilon \mathfrak{G}$.

An endomorphism V of $\mathfrak{G}$ with property (α) is called a projection,[1] and every projection is a splitting endomorphism. Also $\mathfrak{G} = \mathfrak{H}(V) \oplus \mathfrak{K}(V)$, a direct sum of $\perp$ summands. Let b be a fixed element of $\mathfrak{G}$. Define a a mapping P_b on $\mathfrak{G}$ into $\mathfrak{G}$ by $P_b(x) = x - x_{\perp b}$. For brevity, we set $x_b = x - x^{\perp}{}_b$. $x_b \leq x$. $x_b = 0$ if, and only if, $x = x_{\perp b}$, that is, if, and only if, $x \perp b$. Hence $\mathfrak{K}(P_b) = \mathfrak{K}_b$. $x_b \perp x_{\perp b}$ and $x_b \varepsilon \mathfrak{H}_b$. Thus, $(x_b)_{\perp b} = 0$, and $P^2{}_b = P_b$. Also $P_b(x) = x$ if, and only if, $x_{\perp b} = 0$. This shows that $\mathfrak{H}(P_b) = \mathfrak{H}_b$. Now $x = x_b + x_{\perp b}$. Hence $\mathfrak{G} = \mathfrak{H}_b \oplus \mathfrak{K}_b$. Thus, P_b is an endomorphism and therefore a projection. It is onto $\mathfrak{H}_b$, its invariant set. There is a companion projection P'_b defined by $P'_b(x) = x_{\perp b}$. $\mathfrak{H}(P'_b) = \mathfrak{K}_b$ and $\mathfrak{K}(P'_b) = \mathfrak{H}_b$.

The following eight conditions are equivalent. The proofs are elementary. (1) $a \parallel b$. (2) $\mathfrak{H}_a = \mathfrak{H}_b$. (3) $\mathfrak{K}_a = \mathfrak{K}_b$. (4) a is maximal in $\mathfrak{H}_b$. (5) b is maximal in $\mathfrak{H}_a$. (6) $P_a = P_b$. (7) $a \varepsilon \mathfrak{H}_b$ and $b \varepsilon \mathfrak{H}_a$. (8) $a - b \varepsilon \mathfrak{H}_a \cap \mathfrak{H}_b$.

The convexity of $\mathfrak{H}_b$ shows that $x \varepsilon \mathfrak{H}_b$ if, and only if, there exists $x^* \varepsilon \mathfrak{G}$ with $x^* \geq x$, $x^* \parallel b$. In fact, for $x \varepsilon \mathfrak{H}_b$, $x + P'_x(b)$ is an appropriate x^*. x_b is maximum with respect to being in $\mathfrak{H}_b \cap [y \mid y \varepsilon \mathfrak{G};\, y \leq x]$. Note that a is maximal in $\mathfrak{H}_a$, and that $\mathfrak{H}_a$ is the cross-cut of all $\mathfrak{K}_x$ where $x \perp a$. $x \perp a$, $a \parallel b$ imply $x \perp b$. It follows immediately that $\parallel$ is an equivalence relation $\mathfrak{G}$. Under $\parallel$, all maximal elements of $\mathfrak{G}$ form an equivalence class. The following are also elementary: (a) $b_a, a_b \varepsilon \mathfrak{H}_a \cap \mathfrak{H}_b$. (b) $a_b \parallel b_a$. (c) $a + b - b_a \parallel a + b - a_b$, $(a + P'_a(b) \parallel b + P'_b(a))$. (d) $a + b \geq a_b + b_a, a_{\perp b} + b_{\perp a}$, and the latter two sums are $\perp$. In fact, $\mathfrak{H}_a \cap \mathfrak{H}_b \perp \mathfrak{K}_a + \mathfrak{K}_b$ (in the sense of addition of complexes), and the $+$ and $\cap$ can be exchanged in this statement. (e) $a \parallel b$, $a \leq a'$, $b \leq b'$ imply $a + b \leq a' + b'$. (f) If $\mathfrak{G}$ is free of elements of order 2, then $c \geq a, b, a + b$ imply $a \perp b$. (g) $a_b = b_a$ if, and only if, there exists

---

[1] For similar notions in Hilbert space, see Béla v. Sz. Nagy, "Spektraldarstellung linearer Transformationen," Berlin, 1942.

$c \geq a, b$. Thus $c \geq a, b$ and $a \parallel b$ imply $a = b$. (h) If a subset of $\mathfrak{H}_a$ has a l. u. b. in $\mathfrak{G}$, then this l. u. b. lies in $\mathfrak{H}_a$. (*Cf.* section 5, (s^*).) (i) $a \parallel (\perp) b$ if, and only if $a_c \parallel (\perp) b_c$ for every $c \in \mathfrak{G}$. (j) $a + b \geq c$ implies $c = a_c + b_c$. (k) If y is maximal in $\mathfrak{G}$, then $y_x \parallel x$ for every $x \in \mathfrak{G}$.

THEOREM 1. *Let P be a projection of a* v. o. *group $\mathfrak{G}$. Then $P = P_a$ if, and only if, a is maximal in $\mathfrak{H}(P)$; and for each projection P, there exists $a \in \mathfrak{H}(P)$ such that $P = P_a$.*

Proof. Suppose that a is maximal in $\mathfrak{H}(P)$, that $y \perp a$ and that h is the component of y in $\mathfrak{H}(P)$. Then $h \perp a$, and $h + a \geq a$. But $h + a \in \mathfrak{H}(P)$ so that $h = 0$ by the maximal character of a. Hence, $y \in \mathfrak{K}(P)$, and $\mathfrak{K}_a \subset \mathfrak{K}(P)$. $\mathfrak{K}(P) \perp \mathfrak{H}(P)$, and $a \in \mathfrak{H}(P)$ imply $\mathfrak{K}(P) \subset \mathfrak{K}_a$, so that $\mathfrak{K}_a = \mathfrak{K}(P)$. For $x \in \mathfrak{G}$, $x = P(x) + P'(x)$, where $P'(x) \in \mathfrak{K}(P) = \mathfrak{K}_a$. Therefore, $x_a = P_a(x) = P_a P(x)$, and $x_a \leq P(x)$. But $PP_a(x) = P(x_a) = P(x) - P(x_{\perp a}) = P(x)$, since $x_{\perp a} \in \mathfrak{K}_a = \mathfrak{K}(P)$. Thus $P(x) \leq P_a(x) = x_a$; and $P(x) = P_a(x)$ for every $x \in \mathfrak{G}$.. Thus $P = P_a$ if a is maximal in $\mathfrak{H}(P)$. If $P = P_a$, then $\mathfrak{H}(P) = \mathfrak{H}_a$. Since a is maximal in $\mathfrak{H}_a$, a is maximal in $\mathfrak{H}(P)$.

Let b be maximal in $\mathfrak{G}$ and let a be its component in $\mathfrak{H}(P)$. If a is not maximal in $\mathfrak{H}(P)$, then there exist $c \in \mathfrak{H}(P)$, $c \neq a$, $c \geq a$. $b - a \in \mathfrak{K}(P)$, and $\mathfrak{K}(P) \perp \mathfrak{H}(P)$. Hence $c - a \perp b - a$. Also $c - a \perp a$. Thus $c - a \perp (b - a) + a$, or $c - a \perp b$; and $c - a + b \geq b$. Since $c - a \neq 0$, this contradicts the maximal character of b. This shows that a is maximal in $\mathfrak{H}(P)$ so that $P = P_a$, completing the proof.

Now the projections have been put into one-to-one correspondence with the equivalence classes of $\mathfrak{G}$ under $\parallel$, we shall have little difficulty in exhibiting the structure of the set $\mathfrak{P}(\mathfrak{G})$ of projections. We can say that $P_a \leq P_b$ if $P_a(x) \leq P_b(x)$ for every $x \in \mathfrak{G}$. This partly orders $\mathfrak{P}(\mathfrak{G})$. For instance, if $a \geq b$, then $P_a \geq P_b$. More generally, if S and T are endomorphisms of $\mathfrak{G}$, we can say that $S \leq T$ if $S(x) \leq T(x)$ for every $x \in \mathfrak{G}$. P_0 is a universal infimum, which we can denote by 0 without confusion; and $I = P_x$, x maximal in $\mathfrak{G}$, is the identity mapping on $\mathfrak{G}$ and is the universal supremum of $\mathfrak{P}(\mathfrak{G})$. $0 = P_0$ carries all elements of $\mathfrak{G}$ onto 0. In the ordered system $\mathfrak{P}(\mathfrak{G})$ the concepts $\perp$ and $\parallel$ are meaningful and will be used. Products of projections are again projections and are written to the left. That is, $P_a P_b(x) = P_a[P_b(x)]$. Though $P_a + P_b$ is an endomorphism, it need not be a projection. $(x_{\perp a})_{\perp b} \leq x_{\perp b}$. Applying P'_a to both sides, $(x_{\perp a})_{\perp b} \leq (x_{\perp b})_{\perp a}$. This leads to $P'_a P'_b = P'_b P'_a$. If P_r is a projection then there exists $t \in \mathfrak{G}$ such that $P_r = P'_t$. Hence $P_r P_s = P_s P_r$.

The following are equivalent: (1)′ $P_a \perp P_b$ in $\mathfrak{P}(\mathfrak{G})$. (2)′ $a \perp b$.

(3)′ $P_a(x) \perp P_b(x)$ for every $x \varepsilon \mathfrak{G}$. (4)′ $\mathfrak{H}_a \cap \mathfrak{H}_b = (0)$. (5)′ $P_aP_b = 0$. From any one of these it follows that $P_{a+b} = P_a + P_b = P_a \cup P_b$, the l. u. b. of P_a and P_b. If $\mathfrak{G}$ has no elements of order 2, then the following can be added to the equivalence list: (6)′ $P_a + P_b$ is a splitting endomorphism of $\mathfrak{G}$. (7)′ $P_a + P_b \leq I$.

We give another set of equivalent conditions: (1)″ $P_a \geq P_b$. (2)″ $P_b = P_aP_b$. (3)″ $b \varepsilon \mathfrak{H}_a$. (4)″ $\mathfrak{H}_b \subset \mathfrak{H}_a$. (5)″ $\mathfrak{K}_b \supset \mathfrak{K}_a$. (6)″ There exists $c \varepsilon \mathfrak{G}$, $c \parallel b$, $c \leq a$. (7)″ There exists $d \varepsilon \mathfrak{G}$, $d \parallel a$, $d \geq b$.

Immediate are (l) $P_a \leq P_b$ and $c_a \leq d \leq c_b$ imply $d_a = c_a$. (m) If $\mathfrak{H}$ and $\mathfrak{K}$ are subgroups, if $\mathfrak{H} \perp \mathfrak{K}$ and if $\mathfrak{G} = \mathfrak{H} \oplus \mathfrak{K}$, a direct sum, then there exists a projection P_a of $\mathfrak{G}$ such that $\mathfrak{H}_a = \mathfrak{H}$ and $\mathfrak{K}_a = \mathfrak{K}$.

THEOREM 2. *Let $\mathfrak{G}$ be a* v. o. *group. Then the $\mathfrak{H}_a$ and $\mathfrak{K}_a$ are* v. o. *groups, strong if $\mathfrak{G}$ is strong, under the order relation hereditary from $\mathfrak{G}$.*

Proof. Verifying the axioms one by one, we see that axiom (1) is trivial. (2) For $x, y \varepsilon \mathfrak{H}_a$, $x \geq y$, we have $y \perp x - y$ in $\mathfrak{G}$. Since $x \geq y$, $x - y$, with $x \varepsilon \mathfrak{H}_a$, $x \perp x - y$ in $\mathfrak{H}_a$. (3) If $x \perp y$ in $\mathfrak{H}_a$, the convexity of $\mathfrak{H}_a$ shows that $x \perp y$ in $\mathfrak{G}$ so that $x + y \geq x, y$ in $\mathfrak{G}$, and the latter also hold in $\mathfrak{H}_a$. (4) If $x + y \geq x, y$ in $\mathfrak{H}_a$, then the same relations hold in $\mathfrak{G}$ so that $x + y = x \cup y$ in $\mathfrak{G}$. Hence $x + y = x \cup y$ in $\mathfrak{H}_a$. (5) For $c \varepsilon \mathfrak{H}_a$ let $\mathfrak{D} = [x \mid x \varepsilon \mathfrak{H}_a; x \perp c$ in $\mathfrak{H}_a]$. Since the relation $\perp$ in $\mathfrak{H}_a$ implies the relation $\perp$ in $\mathfrak{G}$, axiom (5) shows that $\mathfrak{D}$ is a subgroup of $\mathfrak{H}_a$. Let $\mathfrak{E} = [x \mid x \varepsilon \mathfrak{H}_a; z \leq x$, $z \perp c$ in $\mathfrak{H}_a$ imply $z = 0]$. It turns out that $\mathfrak{E} = \mathfrak{H}_a \cap \mathfrak{H}_c$, so that $\mathfrak{E}$ is a subgroup of $\mathfrak{H}_a$. (6) (a) For $x \varepsilon \mathfrak{H}_a$, there exists $y \varepsilon \mathfrak{G}$, y maximal, $y \geq x$. $P_a(y) \geq x$, and $P_a(y)$ is maximal in $\mathfrak{H}_a$. (6) (b) and (c). If $b, c \varepsilon \mathfrak{H}_a$, the elements $b_{\perp c}$ and bc are in $\mathfrak{H}_a$. (6′) If $\mathfrak{G}$ is strong, so is $\mathfrak{H}_a$, by (h) above. The same statements are true for $\mathfrak{K}_a$, since every $\mathfrak{K}_a$ is an $\mathfrak{H}_b$, by (m) above.

7. $\mathfrak{P}(\mathfrak{G})$ as a Boolean algebra. Let us order the projections as above. Since $b_a \perp b_{\perp a}$, $P_b = P_c + P_{c'}$ where $c = b_a$, $c' = b_{\perp a}$. $P_aP_b = P_aP_c + P_aP_{c'} = P_aP_c + 0 = P_c$, by (2)″ and (3)″ of section 6. Since $c'' = a_b \parallel b_a$, by (b) of section 6, $P_aP_b = P_c = P_{c''}$ by (1) and (6) of section 6. $P_c = P_{c''} \leq P_a$, P_b by (1)″ and (3)″ of section 6. If $P_x \leq P_a, P_b$, then $x \varepsilon \mathfrak{H}_a \cap \mathfrak{H}_b$. By (6)″, $x \parallel t$ where $t \leq a$. Then $x_b \parallel t_b$, $t_b \leq a_b = c''$. Since $x \varepsilon \mathfrak{H}_b$ implies $x = x_b$, $P_x \leq P_{c''} = P_c$. This establishes P_c as $P_a \cap P_b$, the greatest lower bound (g. l. b.) of P_a and P_b. Note that on $\mathfrak{H}_b$ the mappings P_a and P_c coincide, if $c = b_a$.

If we let $c = a + b - b_a = a + b_{\perp a}$, or if we let $c' = a + b - a_b = b + a_{\perp b}$, (2)′ and (c) of section 6 show that $P_c = P_{c'} = P_a + P_r = P_b + P_s$, where $r = b_{\perp a}$ and $s = a_{\perp b}$. It is easy to prove that $P_c = P_a \cup P_b$.

$\mathfrak{P}(\mathfrak{G})$ now appears to be a lattice with universal supremum and infimum under its order relation $\leq$. Ready results are $P_a \cup P'_a = I$, $P_a \cap P'_a = 0$, so that the lattice is complemented. Moreover, $(P_a \cup P_b) \cap (P_a \cup P_c) = (P_a + P_d)(P_a + P_e)$, where $d = b_{\perp a}$ and $e = c_{\perp a}$. This product reduces to $P_a + P_f$, where $f = P_e(d)$, using $(2)'$ and $(5)'$ of section 6. Let $g = b_c$, $h = g_{\perp a}$. Then $P_a \cup (P_b \cap P_c) = P_a \cup P_g = P_a + P_h$. Now $g = P_c(b)$, $b = b_a + d$ and $c = p + e$, where $p = c_a$, $p \perp e$. Then $g = P_p(b_a) + f$, after appropriate decomposition, distribution and reduction. It follows that $h = g_{\perp a} = f_{\perp a} = f$, so that the distributive law $P_a \cup (P_b \cap P_c) = (P_a \cup P_b) \cap (P_a \cup P_c)$ is established. This proves

THEOREM 3. *Let* $\mathfrak{G}$ *be a* v. o. *group. Under the ordering relation* $\leq$ *(as described above),* $\mathfrak{P}(\mathfrak{G})$, *the set of projections of* $\mathfrak{G}$, *is a Boolean algebra.*

8. The complete algebra.

In this section, we shall prove that strong $\mathfrak{G}$ leads to complete $\mathfrak{P}(\mathfrak{G})$. Even if $\mathfrak{G}$ is not strong, projections preserve l. u. b.'s of subsets of $\mathfrak{G}$, whenever these exist. For, if b is the l. u. b. of the elements of a subset $\mathfrak{B}$ of $\mathfrak{G}$, if P_a is a projection and if c is any upper bound for the elements of the set $P_a(\mathfrak{B})$, then $c_a + b_{\perp a}$ is an upper bound for the elements of $\mathfrak{B}$, by (k) of section 5. Since $c \geq c_a$ and since $c_a + b_{\perp a} \geq b$, $c_a \geq b_a$ and $c \geq c_a \geq b_a$. It follows that $b_a = P_a(b)$ is the required l. u. b. of $P_a(\mathfrak{B})$.

THEOREM 4. *Let* $\mathfrak{G}$ *be a* v. o. *group, and let* P *be the* l. u. b. *in* $\mathfrak{P}(\mathfrak{G})$ *of a class* $\mathfrak{R}$ *of projections.* (a) *If, for fixed* x *and* z, *all* $R(x) \leq z$, $R \in \mathfrak{R}$, *then* $P(x) \in \mathfrak{H}_z$. (b) *For each* $x \in \mathfrak{G}$, $P(x)$ *is minimal with respect to being an upper bound for all* $R(x)$. (c) *If, for fixed* x, l. u. b. $R(x)$ *exists, then it equals* $P(x)$.

Proof. (a) If $z = z(y) \geq R(y)$ for every $R \in \mathfrak{R}$, y fixed in $\mathfrak{G}$, then $y \geq P(y) \geq d \geq R(y)$ for every R, where $d = P_zP(y)$. Form an endomorphism $S = P_dP_y + P - PP_y$. $S(y) = d$. $S \leq P$, and S is a projection. $S(x_{\perp y}) = P(x_{\perp y}) \geq R(x_{\perp y})$ for every R. $S(x_y) = P_d(x_y)$. Let B^* denote the projection B restricted to $\mathfrak{H}_y$. Let $g = R^*(y)$ for some fixed $R \in \mathfrak{R}$; and let $h = P_y(c)$ where c is a maximal element of $\mathfrak{H}(R)$. $g, h \in \mathfrak{H}_y$. $R^* = P^*_h$, by a remark in the proof of Theorem 3. By (k) of section 6, $g \parallel h$ in $\mathfrak{H}_y$ so that $P^*_g = P^*_h$. Since $g \leq d$, $P^*_h \leq P^*_d$, and $P^*_h(x_y) \leq P^*_d(x_y) = S^*(x_y)$. But $P^*_h(x_y) = P^*_yR^*(x_y) = R^*(x_y)$. Hence $R(x_y) \leq S(x_y)$. Now $x_{\perp y} \perp x_y$, and S is a projection, so that $S(x_{\perp y}) \perp S(x_y)$. By (k) of section 5, $S(x) \geq R(x)$ for every $x \in \mathfrak{G}$. This shows that $S \geq R$ for every $R \in \mathfrak{R}$; and $S \geq P$. Since $S \leq P$, we have $S = P$. $S(y) = d$ implies $P(y) = d = P_zP(y)$, and

$P(y) \in \mathfrak{H}_z$, proving (a). The proof of (b) is immediate, and (c) follows from (b).

A similar result states that if c is the l. u. b. of a class $\mathfrak{C}$ of elements of $\mathfrak{G}$, a v. o. group, then the class $\mathfrak{P}(\mathfrak{C})$ of projections P_x, $x \in \mathfrak{C}$, has l. u. b. P_c. To prove this, we show that P_c is minimal with respect to being an upper bound for $\mathfrak{P}(\mathfrak{C})$. Since $\mathfrak{P}(\mathfrak{G})$ is a lattice, P_c is the l. u. b. for $\mathfrak{P}(\mathfrak{C})$.

We shall use the statement (e) of section 6 to prove some of the following results. Two subsets $\mathfrak{A}$ and $\mathfrak{B}$ of an ordered system $\mathfrak{G}$ are said to be order isomorphic if there is a one-to-one map f on $\mathfrak{A}$ onto $\mathfrak{B}$ such that both f and f^{-1} preserve order.

LEMMA 2. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be order isomorphic chains of a* v. o. *group $\mathfrak{G}$. Let $\mathfrak{C}$ be the set of all $a_b + b_a$ where $a \in \mathfrak{A}$, $b \in \mathfrak{B}$ and a and b correspond under the order isomorphism f. Then $\mathfrak{C}$ is a chain, and there exists a (many-one) order preserving map g on $\mathfrak{A}$ onto $\mathfrak{C}$ (an order homomorphism). If $\mathfrak{A}$ and $\mathfrak{B}$ have upper bounds a^* and b^*, respectively, then $\mathfrak{C}$ has an upper bound $a^* + b^*$.*

Proof. Since $a_b \parallel b_a$, (e) of section 6 shows that $a_b \leq a'$, $b_a \leq b'$ imply $a_b + b_a \leq a' + b'$. Suppose that $a \leq r, a, r \in \mathfrak{A}$. By hypothesis, $f(a) = b \leq f(r)$. Denote $f(r)$ by s. $a_b \leq a_s \leq r_s$. Let $b_a \leq s_r$. Taking $a' = b_s$, $b' = s_r$, we have $a_b + b_a \leq r_s + s_r$. Define the mapping g by $g(a) = a_b + b_a$ where $f(a) = b$. The above shows that $a \leq r$ implies $g(a) \leq g(b)$. The remaining statements of the lemma follow directly.

LEMMA 3. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be order isomorphic chains of a* v. o. *group $\mathfrak{G}$. Let $\mathfrak{C}$ be the set of all a_b where $a \in \mathfrak{A}$, $b \in \mathfrak{B}$ and a and b correspond under the isomorphism: $b = f(a)$. Then $\mathfrak{C}$ is a chain and is an order homomorph of $\mathfrak{A}$ under a mapping g. Let a^* and b^* be the respective* l. u. b.'*s of $\mathfrak{A}$ and $\mathfrak{B}$. If w is any upper bound for $\mathfrak{C}$, then $a^*_{b^*} \in \mathfrak{H}_w$. $a^*_{b^*}$ is an upper bound for $\mathfrak{C}$ and is its* l. u. b. *if $\mathfrak{C}$ has a* l. u. b. *c^*.*

Proof. That $\mathfrak{C}$ is a chain, that g exists with $g(a) = a_b$, $b = f(a)$, and that $a^*_{b^*}$ is an upper bound for $\mathfrak{C}$ are all evident. Suppose that w is an upper bound for all $a_b \in \mathfrak{C}$. Since $a^* =$ l. u. b. a, $a \in \mathfrak{A}$, and since projections preserve l. u. b.'s, $a^*_b =$ l. u. b. a_b, b fixed in $\mathfrak{B}$. Hence $w \geq a^*_b$ for all $b \in \mathfrak{B}$. Now, $b^* =$ l. u. b. b, $b \in \mathfrak{B}$ implies $P_{b^*} =$ l. u. b. P_b. By Theorem 4(a), $P_{b^*}(a^*) = a^*_{b^*} \in \mathfrak{H}_w$. If $\mathfrak{C}$ has l. u. b. c^*, then $a^*_{b^*} \geq c^*$ and $a^*_{b^*} \in \mathfrak{H}_{c^*}$. Thus, $P_{c^*}(a^*_{b^*}) = c^* = a^*_{b^*}$.

LEMMA 4. (a) *Let $\mathfrak{A}$ and $\mathfrak{B}$ be subsets of $\mathfrak{G}$, a* v. o. *group, and let $\mathfrak{A}$ and $\mathfrak{B}$ be in one-to-one correspondance via a mapping f. Suppose that $a \parallel f(a)$*

for every $a \varepsilon \mathfrak{A}$ *and that* $a, f(a) \varepsilon \mathfrak{H}_d$, *where* $d = a + f(a)$, $d = d(a)$. *If all three quantities exist,* l. u. b. $a +$ l. u. b. $f(a) =$ l. u. b. $d(a)$. (b) *For* $x, y \varepsilon \mathfrak{G}$ *suppose that* $x \parallel y \parallel x + y$ *and that* $\mathcal{R}$ *is a set of projections on* $\mathfrak{G}$. *If the three quantities exist,* l. u. b. $P(x) +$ l. u. b. $P(y) =$ l. u. b. $P(x + y)$, *where the* l. u. b. *are taken over all* $P \varepsilon \mathcal{R}$. (c) *Let* $\mathfrak{A}$ *and* $\mathfrak{B}$ *be subsets of* v. o. $\mathfrak{G}$, *where* $\mathfrak{A} \perp \mathfrak{B}$. *Let* $\mathfrak{C}$ *be* $\mathfrak{A} + \mathfrak{B}$, *the addition being that of complexes. Let* $\mathfrak{A}$, $\mathfrak{B}$ *and* $\mathfrak{C}$ *have the respective* l. u. b.'s a^*, b^* *and* c^*. *Then* $a^* \perp b^*$, *and* $c^* = a^* + b^*$. (d) *Suppose that* $x \perp y$ *in* $\mathfrak{G}$ *and that* $\mathcal{L}$ *is a chain of projections on* $\mathfrak{G}$. *Suppose that* l. u. b. $P(x)$, l. u. b. $P(y)$ *and l. u. b.* $P(x + y)$ *exist, where the* l. u. b. *are taken over all* $P \varepsilon \mathcal{L}$. *Then* l. u. b. $P(x + y)$ $=$ l. u. b. $P(x) +$ l. u. b. $P(y)$, *and* l. u. b. $P(x) \perp$ l. u. b. $P(y)$.

Proof. (b) is a corollary of (a) and (d) is a corollary of (c). (a) and (c) have uncomplicated proofs.

THEOREM 5. *The set* $\mathcal{P}(\mathfrak{G})$ *of projections of a strong vector ordered group* $\mathfrak{G}$ *forms, under its natural order, a complete Boolean algebra.*

Proof. We have already proved that $\mathcal{P}(\mathfrak{G})$ is a Boolean algebra. Let $\mathcal{A}$ be a subset of $\mathcal{P}(\mathfrak{G})$. P_0 is a lower bound for the elements of $\mathcal{A}$. Let $\mathcal{B}$ be the non-empty set of lower bounds of $\mathcal{A}$ in $\mathcal{P}(\mathfrak{G})$. If every chain in $\mathcal{P}(\mathfrak{G})$ were to have a l. u. b. in $\mathcal{P}(\mathfrak{G})$, then Zorn's Lemma would provide a maximal element P^* of $\mathcal{B}$. It is easy to see that P^* would then be the required g. l. b. for $\mathcal{A}$, and we could complete the proof by complementation.

Hence let $\mathcal{L}$ be a chain of projections. For $x \varepsilon \mathfrak{G}$ define $P^*(x)$ as l. u. b. $P(x)$, the l. u. b. taken over all $P \varepsilon \mathcal{L}$. We shall prove that P^* is a projection of $\mathfrak{G}$ and that $P^* =$ l. u. b. P, $P \varepsilon \mathcal{L}$. It is clear that $P^*(0) = 0$ and that $P^*(-x) = -P^*(x)$. For $x, y \varepsilon \mathfrak{G}$, set $x_y = s$, $y_x = t$. The three quantities $x_{\perp y}$, $y_{\perp s}$ and $s + t$ are mutually $\perp$. By Lemma 4(d), $P^*(x + y)$ $= P^*(x_{\perp y}) + P^*(y_{\perp x}) + P^*(s + t)$. But $s \parallel t$ and $P_{s+t}(s) \parallel P_{s+t}(t)$, where the sum of the two latter quantities is $s + t$. By Lemma 4(b), $P^*(s + t)$ $= P^*P_{s+t}(s) + P^*P_{s+t}(t)$. The triad $x_{\perp y}$, $P_{s+t}(s)$ and $P'_{s+t}(s)$ consists of mutually $\perp$ elements; likewise for $y_{\perp x}$, $P_{s+t}(t)$ and $P'_{s+t}(t)$. Now $x = x_{\perp y}$ $+ P_{s+t}(s) + P'_{s+t}(s)$, and $y = y_{\perp x} + P_{s+t}(t) + P'_{s+t}(t)$. By Lemma 4(d)

$$P^*(x) + P^*(y) = P^*(x_{\perp y}) + P^*(y_{\perp x}) + P^*P_{s+t}(s) + P^*P_{s+t}(t) + P^*P'_{s+t}(s) + P^*P'_{s+t}(t).$$

But $P'_{s+t}(s) + P'_{s+t}(t) = P'_{s+t}(s + t) = 0$. Since $P^*(-z) = -P^*(z)$, $P^*P'_{s+t}(s) + P^*P'_{s+t}(t) = 0$. Combining results, $P^*(x + y) = P^*(x) + P^*(y)$, and P^* is established as an endomorphism. That $P^* =$ l. u. b. P and that

$P^* \leq I$ are quite trivial. Hence P^* is a projection with the required property, and the proof of the theorem is complete.

THEOREM 6. *A strong* v. o. *group* $\mathfrak{G}$ *is upper conditionally complete and lower complete.*

Proof. The lower completeness is the statement of 6(c′) of Lemma 1. Let $\mathcal{A}$ be a set of projections of $\mathfrak{G}$. These have a l. u. b. P^*, by Theorem 5. By Theorem 4(b), $P^*(x)$ is a minimal upper bound for the $P(x)$, $P \varepsilon \mathcal{A}$, x fixed in $\mathfrak{G}$. The set $\mathfrak{M}$ of such minimal upper bounds is non-empty and has a g. l. b. t which is a upper bound for the $P(x)$. By the minimal character of $P^*(x)$, $t = P^*(x)$ so that $P^*(x) =$ l. u. b. $P(x)$. Now let $\mathfrak{C}$ be a class of elements of $\mathfrak{G}$, and let $c^{\#}$ be an upper bound for the elements of $\mathfrak{C}$. Let $P^* =$ l. u. b. P_c, $c \varepsilon \mathfrak{C}$. In the above, let $x = c^{\#}$. Then $P^*(c^{\#}) =$ l. u. b. $P_c(c^{\#})$. But $P_c(c^{\#}) = c$, so that l. u. b. c, $c \varepsilon \mathfrak{C}$ exists and equals $P^*(c^{\#})$.

We have two immediate corollaries.

COROLLARY 1. $[x \,|\, x \varepsilon \mathfrak{G}; x \leq a]$, *(where a is fixed in strong* v. o. $\mathfrak{G}$*), is, under the order relation hereditary from* $\mathfrak{G}$*, a complete Boolean algebra isomorphic to* $[P_x \,|\, x \varepsilon \mathfrak{G}; P_x \leq P_a]$*, where the order relation in the latter set is hereditary from* $\mathcal{P}(\mathfrak{G})$.

COROLLARY 2. *Let* K *be any set covering of strong* v. o. $\mathfrak{G}$*, and let* $\mathcal{S}$ *be any subset of* $\mathfrak{G}$*. Let* g. l. b. $\mathcal{S}$ *have the obvious meaning. Then* g. l. b. $\mathcal{S}$ $=$ g. l. b.(g. l. b. $\mathcal{S} \cap \mathcal{U}$) *where the* $\mathcal{U}$ *are the sets of* K.

Part II.

9. Boolean groups. Let $\mathcal{B}$ be a Boolean algebra with ordering relation $\leq$ and universal supremum I. A partition Π of I is a collection of elements of $\mathcal{B}$ such that (1) $\alpha, \beta \varepsilon \Pi$, $\alpha \neq \beta$ imply $\alpha \perp \beta$, and (2) l. u. b. $\alpha = I$, the l. u. b. taken over all $\alpha \varepsilon \Pi$. We exclude repetitions from a partition and also the element 0, the universal infimum of $\mathcal{B}$. Π' is said to be a refinement of Π, ($\Pi' \prec \Pi$), if to each $\alpha \varepsilon \Pi'$ there exists $\beta(\alpha) \varepsilon \Pi$ with $\alpha \leq \beta(\alpha)$. Given two partitions of I, Π and Π', there exists a common refinement $\Pi'' = [\alpha \cap \beta \,|\, \alpha \varepsilon \Pi, \beta \varepsilon \Pi']$, as the infinite distributive law for Boolean algebras shows [2, p. 165]. Call this the intersection refinement of Π and Π'.

Let Σ be a fixed, non-trivial additive Abelian group (not necessarily ordered in any way). This will be called the group of coefficients. To each $\beta \varepsilon \Pi$, a partition, attach a coefficient $s(\beta) \varepsilon \Sigma$. Designate the new object thus formed by (Π, s), where s is the function $\beta \rightarrow s(\beta)$. (Π, s) is to be called a valued partion (of I in $\mathfrak{B}$). We can say that (Π', s') is a refinement of $(\prec)$ (Π, s) if (1) $\Pi' \prec \Pi$ and (2) $\alpha \varepsilon \Pi$, $\beta \varepsilon \Pi$, $\alpha \leq \beta$ imply $s(\beta) = s'(\alpha)$. We can define $(\Pi, s) \sim (\Pi', s')$ if these two valued partitions have a common refinement. If $(\Pi, s) \sim (\Pi, s')$, then $s = s'$. Given (Π, s) and $\Pi' \prec \Pi$, there exists s' with $(\Pi, s) \sim (\Pi', s')$. The relation $\sim$ is an equivalence relation on the set of valued partitions of I of $\mathfrak{B}$. Symmetry and reflexivity are trivial. As for transitivity, suppose that $(\Pi_1, s_1) \sim (\Pi_2, s_2)$, $(\Pi_2, s_2) \sim (\Pi_3, s_3)$ with respective common refinements (Π_{12}, s_{12}) and (Π_{23}, s_{23}). Let Π be the intersection refinement of Π_{12} and Π_{23}. A given $\alpha \varepsilon \Pi$ equals $\beta_1 \cap \beta_3$, $\beta_1 \varepsilon \Pi_{12}$, $\beta_3 \varepsilon \Pi_{23}$. Also there exist $\gamma_{1i} \varepsilon \Pi_i$, $i = 1, 2$, and $\gamma_{3j} \varepsilon \Pi_j$, $j = 2, 3$, with $\beta_1 \leq \gamma_{11}, \gamma_{12}$ and $\beta_3 \leq \gamma_{32}, \gamma_{33}$. Hence $\alpha \leq \gamma_{12} \cap \gamma_{32}$, so that the latter is not 0. Hence $\gamma_{12} = \gamma_{32}$ since Π_2 can have no repetitions and no overlapping elements. Define a mapping s on Π into Σ by $s(\alpha) = s_2(\gamma_{12})$ $(= s_2(\gamma_{32}))$. Then $(\Pi, s) \prec (\Pi_1, s_1)$, (Π_3, s_3) and the transitivity is established.

We denote the set of equivalence classes under $\sim$ by $\mathfrak{G}(\mathfrak{B}, \Sigma)$ or simply by $\mathfrak{G}$, if no confusion arises. Each equivalence class $g \varepsilon \mathfrak{G}$ can be denoted by $\{(\Pi, s)\}$, and we say that $(\Pi, s) \varepsilon g$, or that g is represented by a valued partition (Π, s) (and by all equivalent valued partitions). We can introduce an addition in $\mathfrak{G}$ quite simply. If $(\Pi_1, s_1) \varepsilon g$, $(\Pi_2, s_2) \varepsilon h$, let Π_{12} be the intersection refinement of Π_1 and Π_2, and define s_{12} on Π_{12} into Σ by $s_{12}(\alpha \cap \beta) = s_1(\alpha) + s_2(\beta)$, $\alpha \varepsilon \Pi_1$, $\beta \varepsilon \Pi_2$. Suppose that $(\Pi_0, s_0) \varepsilon \{(\Pi_1, s_1)\} = g$. From (Π_0, s_0) and (Π_2, s_2) form (Π_{02}, s_{02}) by the above addition process. Construct a common refinement (Π, s) of (Π_0, s_0) and (Π_1, s_1). Apply the above addition process to (Π, s) and (Π_2, s_2), obtaining (Π', s'). It is easily seen that (Π', s') is a common refinement of (Π_{12}, s_{12}) and (Π_{02}, s_{02}), so that addition is independent of the members of the equivalence classes which are used as summands. The associativity of addition in $\mathfrak{G}$ follows from that in Σ; and the 0 and the additive inverses can be defined in the obvious way. No confusion will arise if we still denote by $\mathfrak{G}$ the set $\mathfrak{G}$ with this definition of addition. We have proved that $\mathfrak{G}$ is a group, in fact, an Abelian group. We shall call such groups Boolean groups.

$\mathfrak{G}$, a Boolean group, can be partly ordered as follows. For $g, h \varepsilon \mathfrak{G}$, $g \leq h$ if there exist $(\Pi, s_1) \varepsilon g$, $(\Pi, s_2) \varepsilon h$, with $s_1(\alpha) = 0$ or $s_2(\alpha)$ for every $\alpha \varepsilon \Pi$. We call (Π, s_1) and (Π, s_2) $g \leq h$ representatives. If Π' is a refinement of Π, it follows readily that there exist s'_1 and s'_2 such that $(\Pi', s'_1) \prec (\Pi, s_1)$, $(\Pi', s'_2) \prec (\Pi, s_2)$ and (Π', s'_1) and (Π', s'_2) are $g \leq h$ representatives. The following likewise has an immediate proof. If $\mathfrak{B}$ is a complete Boolean algebra, if $(\Pi_1, s_1) \sim (\Pi_2, s_2)$ and if $\sigma \varepsilon \Sigma$, then l. u. b. α, $\alpha \varepsilon \Pi_1$, $s_1(\alpha) = \sigma$, equals l. u. b. β, $\beta \varepsilon \Pi_2$, $s_2(\beta) = \sigma$.

THEOREM 7. *Under the relation $\leq$, every Boolean group $\mathfrak{G}$ is vector ordered.*

Proof. (a) $g \leq g$, for every $g \in \mathfrak{G}$. (b) If $x \leq y$, $y \leq z$, we can, by the above remarks, choose $x \leq y$ and $y \leq z$ representatives so that the representation (Π, s) for y is the same in both cases. It follows at once that $x \leq z$. (c) If $x \leq y$, $y \leq x$, represent y by (Π, s) in both inequalities. This gives x two representations, $(\Pi, s_1) \in x$, where $x \leq y$, and $(\Pi, s_2) \in x$, where $y \leq x$. Hence, $s_1 = s_2$. But $s_1(\alpha) = 0$ or $s(\alpha)$, and $s(\alpha) = 0$ or $s_2(\alpha)$. Since $s_2(\alpha) = s_1(\alpha)$, $s = s_1$ and $x = y$. This establishes $\mathfrak{G}$ as a partly ordered system.

We verify the axioms for v. o. groups in order. (1) is trivial. (2) Let $x \geq y$ representatives be $(\Pi, s_2) \in x$, $(\Pi, s_1) \in y$. Then $x - y = \{(\Pi, s_2 - s_1)\}$. Since $(s_2 - s_1)(\alpha) = 0$ or $s_2(\alpha)$, for every $\alpha \in \Pi$, $x \geq x - y$. If $z \leq y, x - y$, we can redefine Π and find s_3, s'_1 and s^* such that $(\Pi, s_3) \in z$, $(\Pi, s'_1) \in y$, $(\Pi, s^*) \in x - y$. Also, there are the obvious relations between s_3 and s'_1 and between s_3 and s^*: $s_3(\alpha) \neq 0$ implies $s'_1(\alpha) = s^*(\alpha) = s_3(\alpha)$. With the redefinition of Π, $s_1 = s'_1$, and $s^* = s_2 - s_1$. Hence $s_1(\alpha) \neq 0$, and $x \geq y$ implies $s_2(\alpha) = s_1(\alpha)$. This makes $s^*(\alpha) = 0$, a contradiction. It follows that $s_3(\alpha) = 0$ for every $\alpha \in \Pi$, and $z = 0$. We have proved that if $x \geq y$, then $y \perp x - y$.

(3) Suppose that $(\Pi, s_1) \in x$; $(\Pi, s_2) \in y$, $(\Pi, s_3) \in z$ and that $w \geq x, y$ where $x \perp y$. The standard relations are to hold between s_1 and s_3 and between s_2 and s_3. If $s_1(\beta)$ and $s_2(\beta)$ are both non-zero, then $s_1(\beta) = s_2(\beta) = s_3(\beta)$. Let $s(\alpha) = 0$ if $\alpha \in \Pi$, $\alpha \neq \beta$, and let $s(\beta) = s_3(\beta)$. $z = \{(\Pi, s)\} \neq 0$, and $z \leq x, y$. Since $x \perp y$, $z = 0$, a contradiction, and not both $s_1(\beta)$ and $s_2(\beta)$ can be non-zero. $x + y = \{(\Pi, s_1 + s_2)\}$; but $(s_1 + s_2)(\alpha) = s_1(\alpha)$ if the latter is non-zero, $s_2(\alpha)$ otherwise. This is enough to show that $x, y \leq x + y$ if $x \perp y$.

(4) Represent x, y and w as in (3) where, this time, $x + y$, $w \geq x, y$. As in (3), not both $s_1(\alpha)$ and $s_2(\alpha)$ can be non-zero, and we can arrange things so that $s_1(\alpha)$ and $s_2(\alpha)$ are 0 or $s_1(\alpha) + s_2(\alpha)$. If $s_2(\alpha) \neq 0$, $s_1(\alpha) + s_2(\alpha) = s_2(\alpha) = s_3(\alpha)$. If $s_2(\alpha) = 0$, $s_1(\alpha) + s_2(\alpha) = s_1(\alpha)$ which is 0 or $s_3(\alpha)$. This shows that $x + y \leq w$. Therefore, if $x + y \geq x, y$, then $x + y = x \cup y$.

(5) (a) If $(\Pi, s_1) \in x$, $(\Pi, s) \in a$, $x \perp a$, then $s_1(\alpha)$ and $s(\alpha)$ cannot both be non-zero, as in (3) above. $(\Pi, -s_1) \in -x$, where $-s_1$ is defined in the obvious way. $(\Pi, s - s_1) \in a - x$. But $(s - s_1)(\alpha) = s(\alpha)$ if the latter is non-zero, $-s_1(\alpha)$ otherwise. Therefore $a \leq a - x$. By (2) above, $a \perp -x$.

If $x, y \perp a$, a similar procedure shows that $a \leq a + x + y$. Applying (2), $a \perp x + y$. Hence $\mathfrak{K}_a$ is a subgroup of $\mathfrak{G}$.

(5)(b) We can use (2), (5)(a) and (3) to show that $x \varepsilon \mathfrak{H}_a$ implies $-x \varepsilon \mathfrak{H}_a$. Write, now, $(\Pi, s_1) \varepsilon x$, $(\Pi, s_2) \varepsilon y$, $(\Pi, s) \varepsilon a$, $(\Pi, s') \varepsilon u$ where $w \leq x + y$, $u \perp a$ and where $x, y \varepsilon \mathfrak{H}_a$. As above, $s(\alpha)$ and $s'(\alpha)$ cannot both be non-zero. If $s'(\beta) \neq 0$ for some $\beta \varepsilon \Pi$, then $s(\beta) = 0$ and not both $s_1(\beta)$ and $s_2(\beta)$ are 0, say, $s_1(\beta) \neq 0$. Define t by $t(\alpha) = 0$ if $\alpha \varepsilon \Pi$, $\alpha \neq \beta$, and $t(\beta) = s_1(\beta)$. Let $u' = \{(\Pi, t)\}$. $u' \leq x$, $u' \neq 0$. $t(\beta) = s_1(\beta) \neq 0$, while $s(\beta) = 0$. t and s never assume non-zero values together over Π. Hence $u' + a \geq u', a$, so that, by (2), $u' \perp a$. Since $x \varepsilon \mathfrak{H}_a$, $u' = 0$, a contradiction. Thus $s'(\beta) = 0$ for every $\beta \varepsilon \Pi$, and $u = 0$. This proves that $\mathfrak{H}_a$ is a subgroup of $\mathfrak{G}$.

(6)(a) If $x = \{(\Pi, s)\}$ and if x is not maximal, there exists at least one $\alpha \varepsilon \Pi$ for which $s(\alpha) = 0$. Fix $\sigma \varepsilon \Sigma$, $\sigma \neq 0$. Define t by $t(\beta) = s(\beta)$ if $s(\beta) \neq 0$, $t(\beta) = \sigma$ otherwise. Let $m = \{(\Pi, t)\}$. Then m is maximal in $\mathfrak{G}$ and $m \geq x$.

(6)(b) If $x = \{(\Pi, s_1)\}$, $y = \{(\Pi, s_2)\}$, define s by $s(\alpha) = s_1(\alpha)$ if $s_2(\alpha) = 0$, and $s(\alpha) = 0$ otherwise. If we take $z = \{(\Pi, s)\}$, a direct proof shows that $z = x_{\perp y}$.

(6)(c) For x, y as in (6)(b), define t by $t(\alpha) = s_1(\alpha)$ if $s_1(\alpha) = s_2(\alpha)$, and $t(\alpha) = 0$ otherwise. Let $w = \{(\Pi, t)\}$. Then a simple check shows that $w = xy$, the g. l. b. of x and y.

10. Cardinal restrictions on partitions. In section 9, no restrictions were placed on the number of elements. However, if we demand that the power of the class of elements in a given partition Π is $< \aleph$ (or $\leq \aleph$), a fixed cardinal, none of the proofs of section 9 is materially altered. In particular, we can restrict partions to have only a finite collection of elements.

If Σ is a ring, $\mathfrak{G}(\mathcal{B}, \Sigma)$ will also be a ring if we introduce multiplication into $\mathfrak{G}$ in the same manner as addition was introduced. Since the term Boolean ring is well established for another sort of ring, we shall call the present ring a Boolean type ring. It will turn out that every Boolean ring with a unity is also a Boolean type ring. We return to these rings later.

THEOREM 8. *If* $\mathcal{B}$ *is a complete Boolean algebra, then* $\mathfrak{G}(\mathcal{B}, \Sigma)$ *is a strong* v. o. *group.*

Proof. We make no restriction on the cardinals of the partitions Π. Let $\mathcal{C}$ be a chain of elements of $\mathfrak{G}$. For each non-zero $\sigma \varepsilon \Sigma$, construct $\beta(\sigma)$, the union (l. u. b.) of all elements which are valued with σ in all partitions

Π which are used to represent the elements of $\mathfrak{C}$. Let $\beta(0)$ be the complement in $\mathfrak{B}$ of the union of all the $\beta(\sigma)$. $\beta(0) \cap \beta(\sigma) = 0$ for every $\sigma \varepsilon \Sigma$, $\sigma \neq 0$. Suppose that $\beta(\sigma) \cap \beta(\tau) \neq 0$, where σ and τ are distinct, non-zero elements of Σ. Then there exist two distinct members of the chain $\mathfrak{C}$, say, $x \leq y$, $x = \{(\Pi, s_1)\}$, $y = \{(\Pi, s_2)\}$, and there exists $\alpha \varepsilon \Pi$ with $s_1(\alpha) = \sigma$, $s_2(\alpha) = \tau$. But $x \leq y$ and $s_1(\alpha) = \sigma \neq 0$ imply $s_2(\alpha) = s_1(\alpha)$ or that $\tau = \sigma$, a contradiction. Thus $\beta(\sigma) \cap \beta(\tau) = 0$ for every pair of unequal σ and τ. On the other hand, the l. u. b. of all the $\beta(\sigma)$ is I, the supremum of $\mathfrak{B}$, and the $\beta(\sigma)$ form a partition of I. Call this partition Ω. Define t by $t(\beta(\sigma)) = \sigma$, a mapping on Ω into Σ, and let $x^* = \{(\Omega, t)\}$. It is clear that x^* is the l. u. b. of the elements of $\mathfrak{C}$.

11. The representation theorems. Let $\mathfrak{B}$ be a Boolean algebra and let Σ, the coefficient group, be non-trivial. Fix $\beta \varepsilon \mathfrak{B}$ and agree to use only those partitions of I which (1) have only a finite number of elements and (2) are refinements of the partition $\{\beta, \beta'\}$, where β' is the complement of β in $\mathfrak{B}$. For $x = \{(\Pi, s)\}$, let $x(\beta) = \{(\Pi, s')\}$, where $s'(\alpha) = s(\alpha)$ if $\alpha \leq \beta$ and $s'(\alpha) = 0$ if $\alpha \leq \beta'$, $\alpha \varepsilon \Pi$. It follows readily that the mapping P_β: $x \rightarrow x(\beta)$ is a projection of $\mathfrak{S}(\mathfrak{B}, \Sigma)$. Define a mapping Φ on $\mathfrak{B}$ into $\mathfrak{P}(\mathfrak{S}(\mathfrak{B}, \Sigma))$ by $\Phi(\beta) = P_\beta$. If $P_\alpha = P_\beta$, then $x(\alpha) = x(\beta)$ for every $x \varepsilon \mathfrak{S}$. If $\gamma \leq \alpha \cap \beta'$, then $x(\gamma) \leq x(\alpha), x(\beta')$. Since $x(\beta') \perp x(\beta)$, $x(\gamma) = 0$ for every $x \varepsilon \mathfrak{S}$. But this is impossible unless $\gamma = 0$. Therefore $\alpha \cap \beta' = 0$, $\alpha \leq \beta$ and similarly $\beta \leq \alpha$. This shows that Φ is one-to-one. If $P \varepsilon \mathfrak{P}$, then $P = P_x$, $x \varepsilon \mathfrak{S}$. Let $x = \{(\Pi, s)\}$. Here, by (1), Π has only a finite number of elements. If $x = 0$, $P = P_0 = P_\beta$, where $\beta \varepsilon \mathfrak{B}$, $\beta = 0$, the infinum of $\mathfrak{B}$. If $x \neq 0$, let $\beta = \text{l. u. b. } \alpha$, $\alpha \varepsilon \Pi$, $s(\alpha) \neq 0$. Then $P = P_x = P_\beta$. We have proved that Φ is onto $\mathfrak{P}$. Now $\alpha \cup \beta = (\alpha \cap \beta) \cup (\alpha \cap \beta') \cup (\alpha' \cap \beta)$, the l. u. b. of three disjoint quantities. As endomorphisms of $\mathfrak{S}$,

$$\begin{aligned} \Phi(\alpha \cup \beta) &= \Phi(\alpha \cap \beta) + \Phi(\alpha \cap \beta') + \Phi(\alpha' \cap \beta) \\ &= \Phi(\alpha \cap \beta) \cup \Phi(\alpha \cap \beta') \cup \Phi(\alpha' \cap \beta) \cup \Phi(\alpha \cap \beta) \\ &= \Phi(\alpha) \cup \Phi(\beta). \end{aligned}$$

For, the Φ's are elements of $\mathfrak{P}$, and in $\mathfrak{P}$, $P \perp Q$ implies $P + Q = P \cup Q$. Moreover, $\alpha = (\alpha \cap \beta) \cup (\alpha \cap \beta')$ and $\beta = (\alpha \cap \beta) \cup (\alpha' \cap \beta)$. As for $\cap$, $\Phi(\alpha') = P_{\alpha'} = P'_\alpha$, as defined in section 6, and P'_α is the complement of P_α in $\mathfrak{P}$, by section 7. Thus $\Phi(\alpha \cap \beta) = \Phi((\alpha' \cup \beta')') = P'_{\alpha' \cup \beta'} = [P_{\alpha'} \cup P_{\beta'}]' = P'_{\alpha'} \cap P'_{\beta'} = P_\alpha \cap P_\beta = \Phi(\alpha) \cap \Phi(\beta)$. We have proved

THEOREM 9. *A Boolean algebra* $\mathfrak{B}$ *has a faithful (lattice isomorphic)* *representation as the set of all projections on a Boolean group (therefore on a vector ordered group)* $\mathfrak{S}(\mathfrak{B}, \Sigma)$.

Now let Σ be a ring so that $\mathfrak{S}$ is a Boolean type ring. A projection P of the group structure of $\mathfrak{S}$ will preserve products so that it is a ring endomorphism. In Theorem 9, we can replace the words "Boolean group" by "Boolean type ring." In particular, let Σ be the ring $\mathfrak{J}_2$ of integers modulo 2. Let us define a map θ on $\mathfrak{B}$ into $\mathfrak{S}(\mathfrak{B}, \mathfrak{J}_2)$ such that $\theta(\alpha) = \{(\Pi, s)\}$, where Π is the partition $\{\alpha, \alpha'\}$ and $s(\alpha) = 1_2$, $s(\alpha') = 0_2$. Note that the product in $\mathfrak{S}$ (componentwise multiplication) coincides with the product of axiom (6)(c). We restrict all partitions to those with only a finite number of elements. Then θ turns out to be one-to-one onto $\mathfrak{S}$. $\theta(\alpha \cap \beta) = \theta(\alpha)\theta(\beta)$, where the product on the right is the ring product in $\mathfrak{S}$. Let $+_2$ denote the symmetric difference in $\mathfrak{B}$. Then $\theta(\alpha +_2 \beta) = \theta((\alpha \cap \beta') \cup (\alpha' \cap \beta))$ $= \theta(\alpha \cap \beta') + \theta(\alpha' \cap \beta)$ since $\alpha \cap \beta' \perp \alpha' \cap \beta$. We can add in $0 = 2\theta(\alpha \cap \beta)$ on the right hand side. Then we see that $\theta(\alpha +_2 \beta) = \theta(\alpha) + \theta(\beta)$. We thus have

THEOREM 10. *Every Boolean algebra has a faithful representation as a set of splitting endomorphisms on a Boolean ring with a unity, namely on itself as a Boolean ring, and every Boolean ring with a unity is a Boolean type ring.*

Finally, if $\mathfrak{B}$ is a complete Boolean algebra, we can drop the restriction that each partition Π have only a finite number of elements. Then $\mathfrak{B}$ will be faithfully represented by the set of projections on a strong v. o. group.

WASHINGTON UNIVERSITY IN SAINT LOUIS.

BIBLIOGRAPHY.

[1] G. Birkhoff, "Lattice ordered groups," *Annals of Mathematics*, vol. 43 (1942), pp. 298-331.

[2] ———, *Lattice Theory* (second edition), New York, 1948.

[3] H. Brandt, "Verallgemeinerung des Gruppenbegriffes," *Mathematische Annalen*, vol. 96 (1926), pp. 360-366.

[4] O. Frink, Jr., "Representations of Boolean algebras," *Bulletin of the American Mathematical Society*, vol. 47 (1941), pp. 755-756.

[5] M. H. Stone, "The theory of representations for Boolean algebras," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 37-111.

[6] A. Tarski, "Zur Grundlegung der Boole'schen Algebra I," *Fundamenta Mathematicae*, vol. 24 (1935), pp. 177-198.

DIRECT DECOMPOSITIONS IN LATTICES.*

By L. Aileen Hostinsky.

---

1. **Introduction.** A comprehensive refinement theorem for direct decompositions of operator loops is proved in [1]. In this paper we give a generalization of that theorem by using the results of a previous paper (see [4]) to translate into lattice theory Fitting's method of investigating the refinement theorems. Thus we define direct decompositions by their decomposition endomorphisms and use products of these decomposition endomorphisms as a principal tool. The definition of endomorphism and the postulates imposed upon the lattice considered are those of [4].

Although previously it had been possible to state assertions in the refinement theorems in terms of the lattice only, the hypotheses of many theorems dealt also with other terminology. This fact is true particularly in those cases where it was attempted to use Fitting's method exclusively. Because of the results in [4], this method can now be used in the lattice also. There are, however, major differences in the proofs since no use can be made here of the elements (in the sense of [1]) and since pairs of complementary decomposition endomorphisms δ, ϵ do not have the property that $\delta + \epsilon = 1$, which was much used previously.

Except for a slight modification of the Splitting Hypothesis, our refinement theorems are exactly analogous to those of [1].

Splitting Hypothesis. *If δ, ϵ is a pair of complementary decomposition endomorphisms of $p/0$ and if α is a decomposition endomorphism of $p/0$, then $\alpha\delta\alpha\epsilon\alpha$ is a uniformly splitting endomorphism of $p/0$.*

Special Refinement Theorem. *If the Splitting Hypothesis is satisfied by p and if $p = a \oplus b = d \oplus e$, then there exist direct decompositions*

$$a = a' \oplus a'', \quad b = b' \oplus b'', \quad d = d' \oplus d'', \quad \text{and} \quad e = e' \oplus e''$$

such that

$$a' \oplus b' = b' \oplus d' = d' \oplus e' = e' \oplus a' \quad \text{and}$$
$$a'' \oplus b'' = b'' \oplus e'' = e'' \oplus d'' = d'' \oplus a''.$$

Definition. *The permutation $i \rightarrow i^*$ of the integers from 1 to n is*

---

* Received March 31, 1950.

said to effect an exchange isomorphism of the direct decomposition $p = a(1) \oplus \cdots \oplus a(n)$ *upon the direct decomposition* $p = b(1) \oplus \cdots \oplus b(n)$ *if for every* i

$$p = a(1) \oplus \cdots \oplus a(i-1) \oplus b(i^*) \oplus a(i+1) \oplus \cdots \oplus a(n).$$

GENERAL REFINEMENT THEOREM. *If the splitting hypothesis is satisfied by* p, *then any two direct decompositions of* p *possess exchange-isomorphic refinements.*

The last section of the paper is devoted to showing the relationship between our results and those of Ore and Kurosch, who previously have dealt with the problem of direct decompositions in Dedekind structures. Through translation into terms of homomorphism and decomposition endomorphism, concepts and statements of [7] are made more concise, and essentially his main theorem is seen to be contained in ours.

The problem of this paper was suggested by Professor Reinhold Baer, whose valuable suggestions the author gratefully acknowledges.

For the convenience of the reader the remaining part of this section consists of a statement of the postulates and those definitions and theorems of [4] which are used here.

The algebraic system considered is a complete modular lattice P upon which is imposed one additional postulate. The elements of P are denoted by lower-case italic letters and the partial ordering by $\leqq$. If p and q are elements in P, there exist (uniquely determined) elements pq, the *product* of p and q, and $p+q$, the *sum* of p and q, such that $pq \leqq p \leqq p+q$, $pq \leqq q \leqq p+q$ and such that $b \leqq p \leqq a$ and $b \leqq q \leqq a$ together imply that $b \leqq pq$ and $p+q \leqq a$. In addition to Dedekind's modular law, we assume also the existence of an element 0 in P such that $0 \leqq p$ for every p in P and the following additional property:

POSTULATE (*). *If* a *and* $\{p_\nu\}$, *where* ν *ranges over a set of ordinals, are elements in* P *and if the* p_ν *form an ascending chain, then* $a \sum_\nu p_\nu = \sum_\nu ap_\nu$.

A single-valued transformation η is a *homomorphism* of p/p' in P upon q/q' in Q (a system satisfying the same postulates as P, or P itself) if η maps x/p' for every x satisfying $p' \leqq x \leqq p$ upon $x\eta/q'$, where $q' \leqq x\eta \leqq q$, so that the following requirements hold:

1) $(\sum_x x)\eta = \sum_x (x\eta)$ (additivity).

2) If $q' \leqq z \leqq x\eta$, then there exists an element z' with $p' \leqq z' \leqq x$ such that $z'\eta = z$ (exhaustiveness).

3) If $x\eta = y\eta$, then there exist elements x' and y' ($p' \leq x' \leq p$ and $p' \leq y' \leq p$) such that $x + x' = y + y'$ and $x'\eta = y'\eta = q'$.

The *kernel* $k(\eta)$ of a mapping η of p/p' on q/q' is the sum of all elements which η maps upon q'. If η is a homomorphism of p/p' upon q/q' and if $p' \leq x \leq p$, then

1) $x\eta = q'$ if and only if $x \leq k(\eta)$.

2) $x\eta = y\eta$ if and only if $x + k(\eta) = y + k(\eta)$.

An *isomorphism* of p/p' upon q/q' is a one-to-one, exhaustive, and additive mapping of p/p' upon q/q'. An isomorphism η of $p/0$ upon $q/0$ has the property $p\eta = q$. An *automorphism* of p/p' is an isomorphism of p/p' upon itself. A homomorphism η of p/p' upon q/q' is an isomorphism if and only if $k(\eta) = p'$. Moreover, products of homomorphisms (isomorphisms) are homomorphisms (isomorphisms).

The mapping η is an *endomorphism* of $p/0$ if $p\eta \leq p$ and if η is a homomorphism of $p/0$ upon $p\eta/0$. The kernels $k(\eta^i)$, for $i = 1, 2, \cdots$, of the endomorphism η^i of $p/0$ form an ascending chain $k(\eta) \leq k(\eta^2) \leq \cdots \leq p$, and the *radical* of the endomorphism η, denoted by $r(\eta)$, is defined to be the sum $\sum_{i=1}^{\infty} k(\eta^i)$.

THEOREM A. *If η is an endomorphism of $p/0$, then*

1) *If $x \leq p$, $x \leq r(\eta)$ if and only if $x\eta \leq r(\eta)$.*

2) *η induces an endomorphism of $r(\eta)/0$.*

If η is an endomorphism of $p/0$ and if $x \leq p$ satisfies $x\eta \leq x$, x is *η-admissible.* If η is an endomorphism of $p/0$ and if x is η-admissible, then η induces in $x/0$ an endomorphism of kernel $xk(\eta)$ and of radical $xr(\eta)$.

The element c is a *complement* of the endomorphism η of $p/0$ if $p = r(\eta) + c$, $cr(\eta) = 0$, and $c\eta = c$. These first two conditions state that p is equal to the *direct sum* $r(\eta) \oplus c$. If there exists a complement of the endomorphism η, then η is said to be a *splitting endomorphism* of $p/0$ or to *split* p; and if η induces a splitting endomorphism of $x/0$ for every η-admissible x, η is a *uniform splitting endomorphism* of $p/0$.

Throughout the remainer of these statements from [4] η denotes an endomorphism of $p/0$ for an arbitrary element p in P.

The element x is said to be *η-automorphic* if $x\eta \leq x \leq p$ and if $y\eta \leq y \leq x$ implies $y\eta = y$.

THEOREM B. *If the endomorphism* η *of* $p/0$ *splits* p *uniformly, then the complement of* η *is uniquely determined,* η*-automorphic, and equal to the sum of all* η*-automorphic elements.*

COROLLARY B. *If* η *splits* p *uniformly with complement* $c(\eta)$ *and if* x *is an* η*-admissible element, then* $xc(\eta)$ *is the uniquely determined complement of the endomorphism which* η *induces in* $x/0$.

THEOREM C. *If* η *is an endomorphism of* $p/0$ *and if for every* η*-admissible* x *there exists a positive integer* $h = h(x)$ *such that* $x\eta^h = x\eta^{h+1} = \cdots$ *and such that* $k(\eta^h) = k(\eta^{h+1}) = \cdots = r(\eta)$, *then* $x = xr(\eta) \oplus x\eta^h$, $x\eta^h$ *is a complement of the endomorphism induced in* $x/0$, *and* η *is uniformly splitting.*

2. Direct decompositions and decomposition endomorphisms. The element p is the *direct sum of* a *and* b ($p = a \oplus b$) if $a + b = p$ and $ab = 0$. (Throughout the paper p represents an arbitrary element in P.)

If $p = a \oplus b$ and if $x \leq p$, there exist uniquely determined mappings α and β which we define by the following equations:

$$x\alpha = (b + x)a \quad \text{and} \quad x\beta = (a + x)b.$$

It follows from these equations that, if $y \leq a$, $y\alpha = y$ and $y\beta = 0$. Similarly, if $z \leq b$, $z\alpha = 0$ and $z\beta = z$.

Upon the basis of these definitions and Dedekind's modular law it can be shown without difficulty that the following properties are possessed by α and β:

1) α and β are endomorphisms of $p/0$.

2) $\alpha^2 = \alpha$ and $\beta^2 = \beta$.

3) $\alpha\beta = \beta\alpha = 0$.

4) $k(\alpha) = p\beta = b$, $k(\beta) = p\alpha = a$, and thus

$$p = a \oplus b = p\alpha \oplus p\beta = k(\beta) \oplus k(\alpha).$$

The pair of endomorphisms α, β is termed a *pair of complementary decomposition endomorphisms.* Either member of the pair is called a decomposition endomorphism. Constant use is made of these facts on decomposition endomorphisms without reference. Likewise, whenever α and β are given as decomposition endomorphisms, the notations just used are to be understood. Similarly, for pairs of decomposition endomorphisms δ, ϵ and ρ, σ the following notations are used:

$$p = d \oplus e = r \oplus s$$

$$d = p\delta = k(\epsilon) \qquad e = p\epsilon = k(\delta)$$

$$r = p\rho = k(\sigma) \qquad s = p\sigma = k(\rho)$$

and for $x \leq p$

$$x\delta = (e + x)d \qquad x\epsilon = (d + x)e$$

$$x\rho = (s + x)r \qquad x\sigma = (r + x)s.$$

The element p is the *direct sum of* $s(1), s(2), \cdots, s(n)$ $(p = s(1) \oplus s(2) \oplus \cdots \oplus s(n))$ if, for every i, p is the direct sum of $s(i)$ and of the sum of the $s(j)$ for $j \neq i$. The following property of direct sums (See, for example, [1], p. 66.) is useful:

$$s(1) \oplus s(2) \oplus \cdots \oplus s(n) = [s(1) \oplus \cdots \oplus s(i)] \oplus [s(i+1) \oplus \cdots \oplus s(n)].$$

3. Preparatory lemmas.

In preparation for the proof of the refinement theorems, we deduce the following lemmas.

Lemma 1. *If $p = a \oplus b$ and if $a \leq s \leq p$, then $s = a \oplus sb$.*

Proof. $a(sb) = (ab)s = 0s = 0$, and $s = s(a + b) = a + sb$, by Dedekind's law.

Lemma 2. *If α, β and ρ, σ are pairs of complementary decomposition endomorphisms of $p/0$ and if $\alpha\rho\alpha$ induces an automorphism of $p\alpha/0$, then $p = p\alpha\rho \oplus p\beta$.*

Proof. From the definitions of the decomposition endomorphisms, it follows that $a\alpha\rho\alpha = [b + (a + s)r]a$. Since $\alpha\rho\alpha$ induces an automorphism of $p\alpha/0 = a/0$, $a = a\alpha\rho\alpha = a\rho\alpha$. Thus $a = [b + (a + s)r]a$, and $a \leq b + (a + s)r$. Hence $p = a + b \leq (a + s)r + b \leq p$, and thus

$$p = (a + s)r + b = a\rho + p\beta = p\alpha\rho + p\beta.$$

Moreover, $(p\alpha\rho)(p\beta) \leq p\alpha\rho$ and ρ is an endomorphism of $p/0$, mapping $p\alpha/0$ $(p\alpha \leq p)$ upon $p\alpha\rho/0$ $(p\alpha\rho \leq p\rho)$. Hence there exists an element $x \leq p\alpha$ such that $x = x\alpha$ and $x\rho (= x\alpha\rho) = (p\alpha\rho)(p\beta)$. Since $(p\alpha\rho)(p\beta) \leq p\beta$,

$$0 = [(p\alpha\rho)(p\beta)]\alpha = x\alpha\rho\alpha.$$

Since $\alpha\rho\alpha$ induces an automorphism of $p\alpha/0$ for $p\alpha \geq x$, the equality $x\alpha\rho\alpha = 0$ implies that $x = 0$. Hence $0 = x\rho = (p\alpha\rho)(p\beta)$, and $p = p\alpha\rho \oplus p\beta$.

Lemma 3. *If α, β is a pair of complementary decomposition endo-*

morphisms of $p/0$ *and if the decomposition endomorphism* ρ *of* $p/0$ *satisfies the conditions* $p\rho \leq p\alpha$ *and* $p\beta \leq k(\rho)$, *then* $\rho = \rho\alpha = \alpha\rho$.

Proof. Since $x\rho \leq p\rho \leq p\alpha$ for any $x \leq p$, $(x\rho)\alpha = x\rho$, or $\rho = \rho\alpha$.

It is to be shown next that $k(\alpha\rho) = k(\rho)$. Since α, β and ρ are decomposition endomorphisms, $p = a \oplus b = p\alpha \oplus p\beta = r \oplus s = p\rho \oplus s$. By hypothesis, $p\beta = b \leq k(\rho) = s$, and hence $s\alpha\rho = [s + (b + s)a]r = [s + sa]r = sr = 0$. Thus $k(\rho) \leq k(\alpha\rho)$. Moreover, $b = k(\alpha) \leq k(\alpha\rho)$. Hence

$$\begin{aligned} k(\alpha\rho) &= k(\alpha\rho)[a + b] \\ &= b + ak(\alpha\rho) \quad \text{(by Dedekind's law).} \end{aligned}$$

Since $0 = [ak(\alpha\rho)]\alpha\rho = [ak(\alpha\rho)]\rho$, it follows that $ak(\alpha\rho) \leq k(\rho)$. Thus, by hypothesis, $k(\alpha\rho) = b + ak(\alpha\rho) \leq k(\rho)$. Hence $k(\alpha\rho) = k(\rho)$.

Since $\alpha^2 = \alpha$, for any $x \leq p$, $x\alpha\rho = (x\alpha)\alpha\rho$. It follows then from the fact that the product $\alpha\rho$ of the endomorphisms α and ρ of $p/0$ is an endomorphism of $p/0$ that $x + k(\alpha\rho) = x\alpha + k(\alpha\rho)$. Hence, by the preceding paragraph, $x + k(\rho) = x\alpha + k(\rho)$. As a result of applying ρ to this equality, $x\rho = [x + k(\rho)]\rho = [x\alpha + k(\rho)]\rho = x\alpha\rho$, or $\rho = \alpha\rho$, as required.

COROLLARY 1. *If* α, β *and* δ, ϵ *are pairs of complementary decomposition endomorphisms of* $p/0$, *if* $p\alpha = u \oplus v$, *and if* $\alpha\delta\alpha$ *induces an automorphism of* $u/0$, *then* $p = u\delta \oplus v \oplus p\beta$.

Proof. Since $p = p\alpha \oplus p\beta = u \oplus (v \oplus p\beta)$, there exist complementary decomposition endomorphisms ρ and σ such that $p\rho = u$ and $p\sigma = v \oplus p\beta$. Furthermore, $p\rho = u \leq p\alpha$ and $p\beta \leq v + p\beta = k(\rho)$. Hence, by Lemma 3, $\rho = \rho\alpha = \alpha\rho$. If $x \leq u$, $x\alpha\delta\alpha \leq u\alpha\delta\alpha = u$ and thus $x\rho = x$ and $(x\alpha\delta\alpha)\rho = x\alpha\delta\alpha$. Hence $x\alpha\delta\alpha = x\rho\alpha\delta\alpha\rho = x\rho\delta\rho$, and thus $\alpha\delta\alpha$ and $\rho\delta\rho$ induce the same automorphism of $u/0$. Hence by Lemma 2, $p = p\rho\delta \oplus p\sigma = u\delta \oplus v \oplus p\beta$.

LEMMA 4. *If* α, β *and* δ, ϵ *are pairs of complementary decomposition endomorphisms of* $p/0$, *then*

1) $(\alpha\delta\alpha)(\alpha\epsilon\alpha) = (\alpha\epsilon\alpha)(\alpha\delta\alpha)$, $\delta\alpha\epsilon = \delta\beta\epsilon$, *and* $\alpha\delta\beta = \alpha\epsilon\beta$.

and for $x\alpha\epsilon\alpha \leq x \leq p\alpha$

2) $x = x\alpha\delta\alpha + x\alpha\epsilon\alpha$.

3) $xr(\alpha\delta\alpha\epsilon\alpha) = xr(\alpha\delta\alpha) \oplus xr(\alpha\epsilon\alpha)$.

4) *An automorphism is induced in* $xr(\alpha\delta\alpha)/0$ *by* $\alpha\epsilon\alpha$ *and in* $xr(\alpha\epsilon\alpha)/0$ *by* $\alpha\delta\alpha$.

Proof. The proofs of 1) and (with the use of 2)) 3) follow directly from the proofs of lemmas 1 to 8 of [6] by a change of notation.

Proof of 2).

$$\begin{aligned}
&b + a\{b + [(x+d)e]\} + a\{b + [(x+e)d]\}\\
&= \{b + [(x+d)e]\}\{a+b\} + \{b + [(x+e)d]\}\{a+b\} && \text{(by Dedekind's law)}\\
&= b + (x+d)e + (x+e)d\\
&= b + (x+d)[e + \{(x+e)d\}] && \text{(by Dedekind's law and } (x+e)d \leq x+d)\\
&= b + (x+d)[(x+e)(e+d)] && \text{(by Dedekind's law and } x+e \geq e)\\
&= b + (x+d)(x+e)\\
&= b + x + (x+d)e && \text{(by Dedekind's law and } x+d \geq x)\\
&= [b + x + (x+d)e][a+b]\\
&= b + a[b + x + (x+d)e] && \text{(by Dedekind's law)}\\
&= b + x + a[b + (x+d)e] && \text{(by Dedekind's law and } x \leq a)\\
&= b + x && \text{(by hypothesis).}
\end{aligned}$$

Hence $a[b + x\alpha\epsilon\alpha + x\alpha\delta\alpha] = a[b+x]$. From this equality, Dedekind's law, $x \leq a$, and $x\alpha\epsilon\alpha + x\alpha\delta\alpha \leq a$, it follows that $x\alpha\epsilon\alpha + x\alpha\delta\alpha = x$.

Proof of 4). Since $\{xk[(\alpha\delta\alpha)^i]\}(\alpha\epsilon\alpha)^i = 0$,

$$\{xk[(\alpha\delta\alpha)^i]\}(\alpha\epsilon\alpha)(\alpha\delta\alpha)^i = \{xk[(\alpha\delta\alpha)^i]\}(\alpha\delta\alpha)^i(\alpha\epsilon\alpha) = 0,$$

by 1), and hence $\{xk[(\alpha\delta\alpha)^i]\}(\alpha\epsilon\alpha) \leq k[(\alpha\delta\alpha)^i]$. Also $\{xk[(\alpha\delta\alpha)^i]\}(\alpha\epsilon\alpha) \leq x\alpha\epsilon\alpha \leq x$, and thus, for all i,

$$\{xk[(\alpha\delta\alpha)^i]\}(\alpha\epsilon\alpha) \leq xk[(\alpha\delta\alpha)^i] \leq xr(\alpha\delta\alpha).$$

Hence

$$\begin{aligned}
(1)\qquad \sum_{i=1}^{\infty}[\{xk[(\alpha\delta\alpha)^i]\}\alpha\epsilon\alpha] &= [\sum_{i=1}^{\infty}\{xk[(\alpha\delta\alpha)^i]\}]\alpha\epsilon\alpha\\
&= [x\sum_{i=1}^{\infty}k[(\alpha\delta\alpha)^i]]\alpha\epsilon\alpha \quad \text{(by Postulate (*))}\\
&= [xr(\alpha\delta\alpha)]\alpha\epsilon\alpha \leq xr(\alpha\delta\alpha).
\end{aligned}$$

It is clear that $\{xk[(\alpha\delta\alpha)^i]\}(\alpha\delta\alpha)^i = 0 \leq [xr(\alpha\delta\alpha)]\alpha\epsilon\alpha$. As induction hypothesis, assume that $\{xk[(\alpha\delta\alpha)^i]\}(\alpha\delta\alpha)^j \leq [xr(\alpha\delta\alpha)]\alpha\epsilon\alpha$ for $0 < j \leq i$. Let $t = \{xk[(\alpha\delta\alpha)^i]\}$. Then, since $t \leq t\alpha\delta\alpha + t\alpha\epsilon\alpha$,

$$t(\alpha\delta\alpha)^{j-1} \leq t(\alpha\delta\alpha)^j + t(\alpha\epsilon\alpha)(\alpha\delta\alpha)^{j-1} = t(\alpha\delta\alpha)^j + t(\alpha\delta\alpha)^{j-1}(\alpha\epsilon\alpha).$$

By the induction hypothesis, $t(\alpha\delta\alpha)^j \leq [xr(\alpha\delta\alpha)]\alpha\epsilon\alpha$. Since $t \leq k[(\alpha\delta\alpha)^i]$,

$$t(\alpha\delta\alpha)^{j-1} \leq k[(\alpha\delta\alpha)^i] \leq r(\alpha\delta\alpha),$$

and

$$t(\alpha\delta\alpha)^{j-1} \leq x(\alpha\delta\alpha)^{j-1} \leq x \qquad \text{(by 2))}.$$

Hence $[t(\alpha\delta\alpha)^{j-1}]\alpha\epsilon\alpha \leq [xr(\alpha\delta\alpha)]\alpha\epsilon\alpha$, and thus $t(\alpha\delta\alpha)^{j-1} \leq [xr(\alpha\delta\alpha)]\alpha\epsilon\alpha$. Hence by complete induction $t = xk[(\alpha\delta\alpha)^i] \leq [xr(\alpha\delta\alpha)]\alpha\epsilon\alpha$. This statement is true for all i, and hence

$$\sum_{i=1}^{\infty} xk[(\alpha\delta\alpha)^i] = xr(\alpha\delta\alpha) \leq [xr(\alpha\delta\alpha)]\alpha\epsilon\alpha.$$

This fact together with (1) shows that $[xr(\alpha\delta\alpha)]\alpha\epsilon\alpha = xr(\alpha\delta\alpha)$.

Moreover, $k(\alpha\epsilon\alpha)xr(\alpha\delta\alpha) \leq xr(\alpha\epsilon\alpha)xr(\alpha\delta\alpha) = 0$, by 3). Hence $\alpha\epsilon\alpha$ induces an automorphism of $xr(\alpha\delta\alpha)/0$. That $\alpha\delta\alpha$ induces an automorphism of $xr(\alpha\epsilon\alpha)/0$ can be proved in a like manner.

LEMMA 5. *If α, β and δ, ϵ are pairs of complementary decomposition endomorphisms of $p/0$ and if $\alpha\delta\alpha\epsilon\alpha$ is a uniformly splitting endomorphism of $p/0$ with complement c, then there exists a direct decomposition $p\alpha = u \oplus v$, where $u = c \oplus p\alpha r(\alpha\delta\alpha)$ and $v = p\alpha r(\alpha\epsilon\alpha)$, and this direct decomposition is uniquely determined by the properties:*

1) *$v \leq r(\alpha\epsilon\alpha)$ and*

2) *an automorphism is induced in $u/0$ by $\alpha\epsilon\alpha$ and in $v/0$ by $\alpha\delta\alpha$.*

Proof. It is to be shown first that $u = c \oplus p\alpha r(\alpha\delta\alpha)$ and $v = p\alpha r(\alpha\epsilon\alpha)$ constitute a direct decomposition of $p\alpha = a$, satisfying 1) and 2). Since $k(\alpha) \leq r(\alpha\delta\alpha\epsilon\alpha)$ and since $p = a \oplus k(\alpha)$, it is a consequence of Lemma 1 that $r(\alpha\delta\alpha\epsilon\alpha) = k(\alpha) \oplus ar(\alpha\delta\alpha\epsilon\alpha)$. Since $a\alpha\epsilon\alpha \leq a$, it follows from Lemma 4 that $ar(\alpha\delta\alpha\epsilon\alpha) = ar(\alpha\delta\alpha) \oplus ar(\alpha\epsilon\alpha)$. Thus

$$r(\alpha\delta\alpha\epsilon\alpha) = k(\alpha) \oplus ar(\alpha\delta\alpha) \oplus ar(\alpha\epsilon\alpha).$$

Because of the hypothesis of splitting, $p = r(\alpha\delta\alpha\epsilon\alpha) \oplus c$, and hence

$$p = k(\alpha) \oplus ar(\alpha\delta\alpha) \oplus ar(\alpha\epsilon\alpha) \oplus c.$$

Since $c = c\alpha\delta\alpha\epsilon\alpha \leq a$, the last three summands and hence their sum are part of a. Lemma 1 and $ak(\alpha) = 0$ then imply that

$$p\alpha = c \oplus ar(\alpha\delta\alpha) \oplus ar(\alpha\epsilon\alpha) = u \oplus v.$$

Since $v = ar(\alpha\epsilon\alpha) \leq r(\alpha\epsilon\alpha)$, 1) is satisfied. By Lemma 4, an automorphism is induced in $ar(\alpha\epsilon\alpha)/0 = v/0$ by $\alpha\delta\alpha$. The next step is to prove that $\alpha\epsilon\alpha$ induces an automorphism of $u/0$.

In this regard it is to be shown first that $c\alpha\epsilon\alpha r(\alpha\delta\alpha\epsilon\alpha) = 0$. Let $c\alpha\epsilon\alpha r(\alpha\delta\alpha\epsilon\alpha) = t$. Since $\alpha\epsilon\alpha$ maps $c/0$ upon $c\alpha\epsilon\alpha/0$ and since $t \leq c\alpha\epsilon\alpha$, there exists $s \leq c$ such that $s\alpha\epsilon\alpha = t$. Thus $s\alpha\delta\alpha\epsilon\alpha = s\alpha\epsilon\alpha\delta\alpha$ (by 1) of Lemma 4) $= t\alpha\delta\alpha \leq [ar(\alpha\delta\alpha\epsilon\alpha)]\alpha\delta\alpha$. By Lemma 4,

$$[ar(\alpha\delta\alpha\epsilon\alpha)]\alpha\delta\alpha = [ar(\alpha\delta\alpha) \oplus ar(\alpha\epsilon\alpha)]\alpha\delta\alpha.$$

Moreover, $[ar(\alpha\delta\alpha)]\alpha\delta\alpha \leq ar(\alpha\delta\alpha)$ and, by Lemma 4, $[ar(\alpha\epsilon\alpha)]\alpha\delta\alpha = ar(\alpha\epsilon\alpha)$. Thus

$$[ar(\alpha\delta\alpha\epsilon\alpha)]\alpha\delta\alpha \leq ar(\alpha\delta\alpha\epsilon\alpha) \leq r(\alpha\delta\alpha\epsilon\alpha).$$

Hence, by 1) of Theorem A, $s \leq r(\alpha\delta\alpha\epsilon\alpha)$. Consequently, since $s \leq c$, $s \leq cr(\alpha\delta\alpha\epsilon\alpha) = 0$. Hence $s\alpha\epsilon\alpha = t = c\alpha\epsilon\alpha r(\alpha\delta\alpha\epsilon\alpha) = 0$. Similarly,

$$c\alpha\delta\alpha r(\alpha\delta\alpha\epsilon\alpha) = 0.$$

Since c is a complement of $\alpha\delta\alpha\epsilon\alpha$,

$$c(\alpha\epsilon\alpha)(\alpha\delta\alpha\epsilon\alpha) = c(\alpha\epsilon\alpha)(\alpha\delta\alpha)(\alpha\epsilon\alpha) = c(\alpha\delta\alpha\epsilon\alpha)(\alpha\epsilon\alpha) = c\alpha\epsilon\alpha.$$

Thus, $c\alpha\epsilon\alpha$ is $\alpha\delta\alpha\epsilon\alpha$-admissible, and since $\alpha\delta\alpha\epsilon\alpha$ is uniformly splitting,

$$c\alpha\epsilon\alpha = c(c\alpha\epsilon\alpha) \oplus c\alpha\epsilon\alpha r(\alpha\delta\alpha\epsilon\alpha) = c(c\alpha\epsilon\alpha),$$

by Corollary B. Hence $c\alpha\epsilon\alpha \leq c$ and $c = (c\alpha\epsilon\alpha)\alpha\delta\alpha \leq c\alpha\delta\alpha$. Similarly $c\alpha\delta\alpha \leq c$ and $c \leq c\alpha\epsilon\alpha$. Hence $c = c\alpha\epsilon\alpha$ (and $c = c\alpha\delta\alpha$). Since, by Lemma 4, $\alpha\epsilon\alpha$ induces an automorphism of $ar(\alpha\delta\alpha)/0$, $[ar(\alpha\delta\alpha)]\alpha\epsilon\alpha = ar(\alpha\delta\alpha)$. Hence

$$u\alpha\epsilon\alpha = [c + ar(\alpha\delta\alpha)]\alpha\epsilon\alpha = c + ar(\alpha\delta\alpha) = u.$$

From $uk(\alpha\epsilon\alpha) \leq ar(\alpha\epsilon\alpha) = v$ it follows that $uk(\alpha\epsilon\alpha) \leq uv = 0$. Thus $\alpha\epsilon\alpha$ induces an automorphism of $u/0$, and the proof of the existence is completed.

Consider a direct decomposition $p\alpha = u' \oplus v'$ satisfying conditions 1) and 2). By 1), $v' \leq ar(\alpha\epsilon\alpha) = v$. Hence, by Lemma 1,

$$v = v' \oplus vu'. \tag{2}$$

By 2), $\alpha\epsilon\alpha$ induces an automorphism of $u'/0$. Thus $(\alpha\epsilon\alpha)^i$ induces an automorphism of $u'/0$, and $u'k[(\alpha\epsilon\alpha)^i] = 0$ for all i. Hence $u'r(\alpha\epsilon\alpha) = 0$ and thus $u'v = 0$. From (2) it now follows that $v = v'$.

Since $v = v'$, $a = u \oplus ar(\alpha\epsilon\alpha) = u' \oplus ar(\alpha\epsilon\alpha)$, and, by 2), $u\alpha\epsilon\alpha = u$ and $u'\alpha\epsilon\alpha = u'$. Hence u and u' are both complements of the endomorphism which $\alpha\epsilon\alpha$ induces in $a/0$ $(a \geq a\alpha\epsilon\alpha)$. Consider any x such that $x \leq a$ and

$x\alpha\epsilon\alpha \leq x$. By 2) of Lemma 4, $x\alpha\delta\alpha \leq x$. Hence $x\alpha\delta\alpha\epsilon\alpha = (x\alpha\delta)\alpha\epsilon\alpha \leq x\alpha\epsilon\alpha \leq x$. Thus, since $\alpha\delta\alpha\epsilon\alpha$ is uniformly splitting, $x = xc \oplus xr(\alpha\delta\alpha\epsilon\alpha)$, by Corollary B. By 3) of Lemma 4, $xr(\alpha\delta\alpha\epsilon\alpha) = xr(\alpha\delta\alpha) \oplus xr(\alpha\epsilon\alpha)$, and hence $x = [xc \oplus xr(\alpha\delta\alpha)] \oplus xr(\alpha\epsilon\alpha)$.

It is known that $(xc)\alpha\delta\alpha \leq x\alpha\delta\alpha c\alpha\delta\alpha \leq xc$ and $(xc)\alpha\epsilon\alpha \leq x\alpha\epsilon\alpha c\alpha\epsilon\alpha \leq xc$. Moreover, since xc is a complement,

$$xc = (xc)\alpha\delta\alpha\epsilon\alpha = [(xc)\alpha\delta\alpha]\alpha\epsilon\alpha \leq (xc)\alpha\epsilon\alpha.$$

Thus $(xc)\alpha\epsilon\alpha = xc$. By 4) of Lemma 4, $[xr(\alpha\delta\alpha)]\alpha\epsilon\alpha = xr(\alpha\delta\alpha)$, and thus

$$[xc \oplus xr(\alpha\delta\alpha)]\alpha\epsilon\alpha = xc \oplus xr(\alpha\delta\alpha).$$

Hence $\alpha\epsilon\alpha$ splits x. Thus the endomorphism which $\alpha\epsilon\alpha$ induces in $a/0$ splits a unformily. Hence, by Theorem B, its complement is unique, and $u' = u$. Thus the uniqueness of the direct decomposition of $p\alpha$ has been shown.

LEMMA 6. *If α, β and δ, ϵ are pairs of complementary decomposition endomorphisms of $p/0$ and if the direct decompositions*

$$p\alpha = a' \oplus a'',\ p\beta = b' \oplus b'',\ p\delta = d' \oplus d'',\ \textit{and}\ p\epsilon = e' \oplus e''$$

have the properties that

α induces an isomorphism of $d'/0$ upon $a'/0$ and of $e''/0$ upon $a''/0$,
β induces an isomorphism of $e'/0$ upon $b'/0$ and of $d''/0$ upon $b''/0$,
δ induces an isomorphism of $a'/0$ upon $d'/0$ and of $b''/0$ upon $d''/0$, and
ϵ induces an isomorphism of $b'/0$ upon $e'/0$ and of $a''/0$ upon $e''/0$

then the following equations are satisfied:

$$\begin{array}{ll} p\alpha \oplus b' = p\alpha \oplus e' & p\alpha \oplus b'' = p\alpha \oplus d'' \\ p\beta \oplus a' = p\beta \oplus d' & p\beta \oplus a'' = p\beta \oplus e'' \\ p\delta \oplus e' = p\delta \oplus b' & p\delta \oplus e'' = p\delta \oplus a'' \\ p\epsilon \oplus d' = p\epsilon \oplus a' & p\epsilon \oplus d'' = p\epsilon \oplus b''. \end{array}$$

Proof. Since $\beta\epsilon\beta$ induces an automorphism of $b'/0$,

$$p = b'\epsilon \oplus b'' \oplus p\alpha = e' \oplus b'' \oplus p\alpha,$$

by Corollary 1. Furthermore, $b' = e'\beta = (e' + a)b \leq e' + a = p\alpha \oplus e'$, and thus $p\alpha \oplus b' \leq p\alpha \oplus e'$.

Since $\beta\epsilon\beta$ induces an automorphism of $b''/0$, it follows from the same corollary that $p = p\alpha \oplus b' \oplus b''\delta = p\alpha \oplus b' \oplus d''$. Hence, by Lemma 1 and the fact that $p\alpha \oplus b' \leq p\alpha \oplus e'$,

$$p\alpha \oplus e' = (p\alpha \oplus b') \oplus d''(p\alpha \oplus e'). \tag{3}$$

By Dedekind's law, $(e' + d'')(a + e') = e' + d''(a + e')$ and also $(e' + d'')(a + e') = e' + a(c' + d'')$. Thus $e' + d''(a + e') = e' + a(e' + d'')$, and hence $[e' + d''(a + e')]\beta = [e' + a(e' + d'')]\beta$. Thus $b' + [d''(a + e')]\beta = b'$, and $[d''(a + e')]\beta \leq b'$. Furthermore, $d'(a + e') \leq d''$, and thus $[d''(a + e')]\beta \leq d''\beta = b''$ (by hypothesis). Hence $[d''(a + e')]\beta \leq b'b'' = 0$. Thus $d''(a + e') \leq k(\beta) = p\alpha$, and hence $d''(a + e') \leq \mathrm{d}''p\alpha = 0$. Hence, by (3), $p\alpha \oplus b' = p\alpha \oplus e'$. The other equations of the assertion can be proved in a like manner.

4. Proof of the refinement theorems. To prove the Special Refinement Theorem let α, β and δ, ϵ denote the pairs of complementary decomposition endomorphisms such that $p\alpha = a$, $p\beta = b$, $p\delta = d$, and $p\epsilon = e$. Lemma 5 asserts the existence of uniquely determined elements a', a'', d', and d'' such that

$$a = a' \oplus a'', \quad d = d' \oplus d'', \quad a'' \leq r(\alpha\delta\alpha), \quad d'' \leq r(\delta\alpha\delta),$$

and an automorphism is induced in $a'/0$ by $\alpha\delta\alpha$, in $a''/0$ by $\alpha\epsilon\alpha$, in $d'/0$ by $\delta\alpha\delta$, and in $d''/0$ by $\delta\beta\delta$.

Let the following definitions hold:

$$b' = b(d' \oplus e), \quad b'' = d''\beta, \quad e' = e(a' \oplus b), \quad \text{and} \quad e'' = a''\epsilon.$$

As in [1] (p. 78) it can be shown that

$$b = b' \oplus b'' \quad \text{and} \quad e = e' \oplus e''.$$

It is to be shown next that these decompositions of a, b, d, and e satisfy the hypotheses of Lemma 6. The following facts can be deduced by essentially the same arguments as in [1] (p. 78):

a) α induces an isomorphism of $e''/0$ upon $a''/0$.
b) β induces an isomorphism of $d''/0$ upon $b''/0$.
c) δ induces an isomorphism of $b''/0$ upon $d''/0$.
d) ϵ induces an isomorphism of $a''/0$ upon $e''/0$.
e) $d = a'\delta \oplus d(a'' \oplus b)$.
f) $\delta\alpha\delta$ induces an automorphism of $a'\delta/0$.

It is to be shown next that $[ar(\alpha\delta\alpha)]\delta \leq r(\delta\alpha\delta)$. Since $\{ak[(\alpha\delta\alpha)^i]\}(\alpha\delta\alpha)^i = 0$, $[\{ak(\alpha\delta\alpha)^i]\}\alpha\delta(\delta\alpha\delta)^i = 0$ and thus

$$\{ak[(\alpha\delta\alpha)^i]\}\alpha\delta = \{ak[(\alpha\delta\alpha)^i]\}\delta \leq k[(\delta\alpha\delta)^i] \text{ for all } i.$$

Hence $\sum_{i=1}^{\infty}\{ak[(\alpha\delta\alpha)^i]\}\delta \leq \sum_{i=1}^{\infty} k[(\delta\alpha\delta)^i]$. Thus, by Postulate (*) and the fact that δ is additive,

$$(4) \qquad [ar(\alpha\delta\alpha)]\delta \leq r(\delta\alpha\delta).$$

Moreover,

$$(5) \qquad [d(a'' \oplus b)]\delta\alpha = [d(a'' \oplus b)]\alpha \leq (a'' \oplus b)\alpha = a''.$$

Since $a'' \leq ar(\alpha\delta\alpha)$, $a''\delta \leq [ar(\alpha\delta\alpha)]\delta \leq r(\delta\alpha\delta)$, by (4). Hence by (5), $[d(a'' \oplus b)]\delta\alpha\delta \leq a''\delta \leq r(\delta\alpha\delta)$. Thus, by 1) of Theorem A, $d(a'' \oplus b) \leq r(\delta\alpha\delta)$.

From $d = a'\delta \oplus d(a'' \oplus b)$, $d(a'' \oplus b) \leq dr(\delta\alpha\delta)$, and Lemma 1, it follows that $dr(\delta\alpha\delta) = d(a'' \oplus b) \oplus dr(\delta\alpha\delta)a'\delta$. Since $\delta\alpha\delta$, and thus $(\delta\alpha\delta)^i$, induces an automorphism of $a'\delta/0$, $a'\delta k[(\delta\alpha\delta)^i] = 0$, for all i. Hence $\sum_{i=1}^{\infty} a'\delta k[(\delta\alpha\delta)^i] = a'\delta r(\delta\alpha\delta) = 0$. Thus $dr(\delta\alpha\delta) = d(a'' \oplus b)$. By 4) of Lemma 4, $\delta\beta\delta$ induces an automorphism of $dr(\delta\alpha\delta)/0$, and hence

an automorphism is induced in $d(a'' \oplus b)/0$ by $\delta\beta\delta$.

The decomposition e) of d thus meets the requirements which, by Lemma 5, determine uniquely the decomposition $d = d' \oplus d''$. Hence

$$(6) \qquad d' = a'\delta \quad \text{and} \quad d'' = d(a'' \oplus b).$$

Likewise it is seen that

$$(7) \qquad a' = d'\alpha \quad \text{and} \quad a'' = a(d'' \oplus e).$$

It is a consequence of (6), (7), and the facts that $\delta\alpha\delta$ induces an automorphism of $d'/0$ and that $\alpha\delta\alpha$ induces an automorphism of $a'/0$ that δ induces an isomorphism of $a'/0$ upon $d'/0$ and α induces an isomorphism of $d'/0$ upon $a'/0$.

Moreover, $k(\beta)e' = ae' = a[e(a' \oplus b)] = ea'$ (by Dedekind's law). Furthermore,

$$(8) \qquad (ea')\alpha\delta\alpha = (ea')\delta\alpha \leq (e\delta)\alpha = 0,$$

and since $\alpha\delta\alpha$ induces an automorphism of $a'/0$, the equality (8) implies that

$ea' = 0$. Thus $e'k(\beta) = 0$. Similarly, since $\delta\alpha\delta$ induces an automorphism of $d'/0$, $b'k(\epsilon) = 0$.

Since $d' = a'\delta$, $d'\delta = a'\delta^2 = a'\delta$. Thus, since δ is an endomorphism of $p/0$, $d' + e = a' + e$. Hence $b' = b(d' + e) = b(a' + e)$. Moreover, by the modular law,

$$(a' + e)(a' + b) = a' + b(a' + e) = a' + b'$$

and

$$(a' + e)(a' + b) = a' + e(a' + b) = a' + e'.$$

Thus $(a' + b')\beta = (a' + e')\beta$, and hence $b' = b'\beta = e'\beta$. Since $a' = d'\alpha$, it follows in a similar manner that $b'\epsilon = e'$. Consequently β induces an isomorphism of $e'/0$ upon $b'/0$, and ϵ induces an isomorphism of $b'/0$ upon $e'/0$.

The proof that the decompositions of a, b, d, and e satisfy the hypotheses of Lemma 6 is thus completed. Hence the decompositions satisfy the equations of the conclusion of that lemma, and we use this fact (as in [1] (pp. 79-80) without repeating the arguments here) to complete the proof of the theorem.

With the preceding results now available, the General Refinement Theorem can be deduced from the Special Refinement Theorem by means of a double induction carried out precisely as in [1] (pp. 81-84), where only structural arguments are used. The following two additional lemmas are required to execute this proof:

LEMMA 7. *If* $p = d \oplus a(1) \oplus \cdots \oplus a(n) = d \oplus b$, *then*

1) $d \oplus a(i) = d \oplus b(i)$, *where* $b(i) = [d \oplus a(i)]b$, *and*

2) $b = b(1) \oplus \cdots \oplus b(n)$.

LEMMA 8. *If the Splitting Hypothesis is satisfied by* p, *then it is satisfied by every direct summand of* p.

The proofs of the analogous lemmas in [1] (p. 70 and p. 72) can be used to derive Lemma 7 and, with a simple modification, Lemma 8. The proof of the theorem is not reproduced in our notation.

5. Relationship to theorems of Ore, Kurosch, and Grayev. Previous investigations relating to direct decompositions in Dedekind structures have been made in [3], [5], [6], and [7].

By comparison with [7], it is seen that what are there termed "reductions" are products of decomposition endomorphisms such as $\alpha\delta$ (in the previous notation) and that the "null-factors" are kernels of powers of products of decomposition endomorphisms. Moreover, the condition of the existence of "regular decompositions" can be stated as follows in terms of decomposition endomorphisms: If α and δ are decomposition endomorphisms such that $x\alpha\delta \leq x$, then $x = x(\alpha\delta)^n \oplus xk[(\alpha\delta)^m]$. This splitting condition is much stronger than our splitting hypothesis.

The main theorem of [7] (the assertion of which is that in two different direct decompositions into direct indecomposable summands both sides contain the same number of summands directly similar in pairs) is proved under the assumption of the existence of regular decompositions. It is to be noted here that we restrict ourselves to complete lattices, while Ore's theorem is general. By means of Lemma 4, Theorem C, and a series of computations (here omitted) it can be shown that the Splitting Hypothesis is a consequence of Ore's condition of the existence of regular decompositions. Thus for complete lattices the following obvious consequence of our theorem is easily seen to contain Ore's results:

If the condition of the existence of regular decompositions is satisfied, then two different direct decompositions into direct indecomposable summands are exchange isomorphic.

The "center" in connection with a pair of direct decompositions has been used by Kurosch ([5]; p. 190) and by Baer ([2], p. 524). The refinement theorems of Kurosch and Grayev (Theorem 1 of [6] and Theorem 1 of [3]) are proved under the assumption that the sequence of centers of a pair of direct decompositions attains zero at a finite step, while we have proved the refinement theorems by assuming the Splitting Hypothesis for every pair of complementary decomposition endomorphisms. It is of interest to note that if the n-th center of a pair of decompositions is zero for an integer n, $(\rho\beta\rho\alpha\rho)^{n+1} = 0$ for α, β a pair of complementary decomposition endomorphisms and ρ a decomposition endomorphism. By comparing this condition with the Splitting Hypothesis, it is noted that $\rho\beta\rho\alpha\rho$ may be uniformly splitting and $(\rho\beta\rho\alpha\rho)^{n+1} \neq 0$.

In regard to the relationship between the two results, it can be stated that the Splitting Hypothesis is a weaker hypothesis on all pairs of complementary decomposition endomorphisms, while Kurosch's hypothesis is a stronger condition on a single pair of direct decompositions. Moreover, the conclusion of direct similarity of refinements is weaker than that of exchange isomorphy. Thus for all pairs of decomposition endomorphisms our result is stronger than that of Kurosch. However, since our theorem is stated from the "universal" point of view, it is independent of the "individualistic" theorem of Kurosch.

TEMPLE UNIVERSITY.

BIBLIOGRAPHY.

[1] R. Baer, "Direct decompositions," *Transactions of the American Mathematical Society*, vol. 62 (1947), pp. 62-98.

[2] ———, "Direct decompositions into infinitely many summands," *Transactions of the American Mathematical Society*, vol. 64 (1948), pp. 519-551.

[3] M. Grayev, "Isomorphisms of direct decompositions in Dedekind structures," *Bulletin de l'Académie des Sciences de l'URSS*, Sér. Math., vol. 11 (1947), pp. 33-45.

[4] L. A. Hostinsky, "Endomorphisms of lattices," *Duke Mathematical Journal*, vol. 18 (1951), pp. 331-342.

[5] A. Kurosch, "Isomorphisms of direct decompositions I," *Bulletin de l'Académie des Sciences de l'URSS*, Sér. Math., vol. 7 (1943), pp. 185-202.

[6] ———, "Isomorphisms of direct decompositions II," *ibid.*, vol. 10 (1946), pp. 47-72.

[7] O. Ore, "Direct decompositions," *Duke Mathematical Journal*, vol. 2 (1936), pp. 581-596.

A GENERALIZATION OF A THEOREM OF JACOBSON.*

By I. N. Herstein.

In [2], Theorem 11, Jacobson proved a theorem which at the same time generalizes the theorem of Wedderburn that every finite division ring is commutative and also the result that any Boolean ring is a commutative ring. The theorem to which we refer is, namely: any ring in which $x^{n(x)} = x$, $n(x)$ an integer greater than one, for every x in the ring, is necessarily commutative.

In this paper we prove a generalization of this result. In fact we prove the following: if in a ring $x^n - x$ is in the center for all ring elements x and a *fixed* integer n which is larger than 1, then the ring must be commutative. The proof is essentially composed of three parts: (1) the proof for semi-simple rings, (2) the location of the radical if the ring possesses one, (3) the proof for subdirectly irreducible rings, from which the general case follows.

The case $n = 2$.[1] Although the case for which $n = 2$ need not be singled out in order to make the proof go through in the general case, the proof in this instance is so simple and elementary that we feel it would be of some interest to present it by itself. Moreover, this situation might be considered as a direct generalization of the fact that a Boolean ring is commutative.

So let us assume that R is a ring with center C, and that for all $x \in R$, $x^2 - x \in C$. Suppose that $x, y \in R$. Thus $(x+y)^2 - (x+y) \in C$. Since both $x^2 - x$ and $y^2 - y$ are in C, we immediately obtain that $xy + yx \in C$; so $x(xy+yx) = (xy+yx)x$. Thus $x^2y = yx^2$, and so $x^2 \in C$. Since $x^2 - x$ is also in C we have that $x \in C$, which is the theorem in this particular case.

The theorem for semi-simple rings. Suppose that K is a division ring with center C and that there exists an integer n greater than 1 so that for all x in K, $x^n - x$ is in C. Suppose that $K \neq C$; that is, there is an $x \in K$, $x \notin C$. Hence (1) $x^n - x \in C$. Let c be any element in C; thus (2) $(cx)^n - cx = c^n x^n - cx \in C$. Combining (1) and (2) we obtain that

---

* Received January 17, 1951.

[1] The case $n = 2$ was conjectured, in a conversation, by Dr. Irving Kaplansky.

$(c^n - c)x \varepsilon C$; so if $c^n - c \neq 0$ it would have an inverse in C, and this would put x in C, contrary to supposition. So we must have that for all $c \varepsilon C$, $c^n = c$. Consequently C must be a *finite* field. Consider the sequence $\{c_i\}$ where $c_i = x^{n^i} - x$. All these c_i's are in C, and since C is finite, for some $i \neq j$, $c_i = c_j$. This leads to $x^s = x$ for some appropriate integer s, $s > 1$. So for all $x \varepsilon K$, $x^{n(x)} = x$, whence by Jacobson's theorem K is commutative. Thus we have shown

THEOREM 1. *If K is a division ring with center C, and if $x^n - x$ is in C for all x in R, and $n > 1$, then K is commutative.*

Before proceeding to the case of primitive and semi-simple rings we establish several easy, but essential, lemmas. We assume that R is a ring with center C and that n is an integer larger than 1.

LEMMA 1. *If for all $x \varepsilon R$, $x^n - x \varepsilon C$, then all the nilpotent elements of R are in C.*

Proof. Suppose that $x^t = 0$. Thus $x^{n^t} = 0$. Since $x^{n^t} - x \varepsilon C$ we have that x is in C.

LEMMA 2. *For the R of Lemma 1 all the idempotents are in the center.*

Proof. Suppose that $e^2 = e$. A simple calculation shows that for any $x \varepsilon R$, $(xe - exe)^2 = (ex - exe)^2 = 0$. Hence by Lemma 1 $xe - exe \varepsilon C$; thus $0 = e(xe - exe) = (xe - exe)e = xe - exe$. That is $xe = exe$; similarly $ex = exe$, whence $xe = ex$ and $e \varepsilon C$.

We consider next the situation where R is a primitive ring [3, p. 312] and where $x^n - x \varepsilon C$. Thus R possesses a maximal right-ideal ρ which contains no non-zero two-sided ideal of R. Let $x \varepsilon \rho$. Since $x^n - x \varepsilon C$, $(x^n - x)R$ is a two-sided ideal of R and is contained in ρ. Hence $(x^n - x)R = (0)$. Since R is primitive this forces $x^n = x$. Let $e = x^{n-1}$. A simple check verifies that $e^2 = e$, so e is in C by Lemma 2. Moreover, eR is then a two-sided ideal of R and is contained in ρ, whence $eR = (0)$, so $e = 0$. Now $x = ex = 0$, whence $\rho = (0)$. So (0) is a maximal right ideal in R, which is primitive, so R is a division ring, which by Theorem 1 is commutative; so we have

THEOREM 2. *If R is a primitive ring with center C, and if for all x in R, $x^n - x$ is in C, then R is commutative.*

Any ring, semi-simple in the sense of Jacobson, is isomorphic to a subdirect sum of primitive rings [3, Theorem 28]. So if R is semi-simple and

$x^n - x \in C$ for all $x \in R$, then R is isomorphic to a subdirect sum of primitive rings, each of which being homomorphic images of R inherit the property that $x^n - x$ is in the center. So, by Theorem 2, each of the components in this representation of R as a subdirect sum is commutative, hence R is itself commutative. Thus

THEOREM 3. *If R is a semi-simple ring and if for all x in R, $x^n - x$ is in the center, then R is commutative.*

The location of the commutators and radical. In what follows we assume that R is a ring with center C and enjoying the property that $x^n - x$ is in C for all x in R, with n a fixed integer larger than 1. Let N be the radical of R (that is, the Jacobson radical). Since R/N is semi-simple and in R/N $x^n - x$ is in the center, by Theorem 3, R/N is commutative. So we immediately obtain

THEOREM 4. *For all $x, y \in R$, $xy - yx \in N$.*

Suppose that $x \in R$, and $c \in C$. Thus (1) $x^n - x \in C$, (2) $(cx)^n - cx = c^n x^n - cx \in C$. Combining (1) and (2) we obtain

LEMMA 3. *For all $x \in R$, and $c \in C$, $(c^n - c)x \in C$.*

COROLLARY. *For all $x, y \in R$ and $c \in C$, $(c^n - c)(xy - yx) = 0$.*

Suppose that $a \in N \cap C$. Thus for all $x, y \in R$, $(a^n - a)(xy - yx) = 0$. Since $a \in N$, $1 - a^{n-1}$ (the unit element is purely formal) is not a divisor of zero. Thus $a \cdot (xy - yx) = 0$. Let $b \in N$. Thus $b^n - b \in N \cap C$, whence $(b^n - b)(xy - yx) = 0$. Repeating the argument as in the case of a, we have $b(xy - yx) = 0$ for all x and y in R. In particular $b(by - yb) = 0$. By the analogous argument, $(by - yb)b = 0$. From these two we obtain $b^2 y = byb = yb^2$; that is, $b^2 \in C$. Now, $b^n - b$ is in C, so if n is even, since $b^2 \in C$, we would have $b^n \in C$ from which it would follow that b is in C. On the other hand, if n is odd, $1 - b^{n-1}$ would be in C, and is not a divisor of zero. This combined with $b(b^{n-1} - 1) \in C$ would put b in C. So in any case, $b \in C$. Thus

THEOREM 5. *$N \subset C$.*

Combining Theorems 4 and 5 we obtain

COROLLARY. *For all x, y in R, $xy - yx$ is in C.*

COROLLARY. *For all x, y in R, $(xy - yx)^2 = 0$.*

The theorem for general rings. We recall that a ring is said to be *subdirectly irreducible* if the intersection of its non-zero two-sided ideals is not the zero ideal. [1, p. 764.] The importance of this concept is brought out in the theorem [1, Theorem 2].

THEOREM 6. *Every ring is isomorphic to a subdirect sum of subdirectly irreducible rings.*

Suppose that R is a ring in which $x^n - x \in C$ for all $x \in R$. R is isomorphic to a subdirect sum of subdirectly irreducible rings, each of which, being isomorphic images of R enjoy the property that $x^n - x$ is in the center. To prove our result it is thus sufficient to prove it for subdirectly irreducible rings. So we suppose that R is subdirectly irreducible and that for all $x \in R$, $x^n - x \in C$, the center of R. Let S be the intersection of the non-zero two-sided ideals of R. $S \neq (0)$.

THEOREM 7. *There exists a prime p so that for all $x, y \in R$, $p(xy - yx) = 0$.*

Proof. Since $x^n - x \in C$ and $(2x)^n - 2x \in C$, we easily obtain that for all $x \in R$, $(2^n - 2)x \in C$, and so for all $x, y \in R$, $(2^n - 2)(xy - yx) = 0$. So if R is not commutative it has non-zero elements of non-zero finite additive order. Let x, y be such that $xy - yx$ has minimal non-zero order p; clearly p is a prime. Let $R_p = \{x \in R \mid px = 0\}$. R_p is a two sided ideal of R, and is not (0), hence $R_p \supset S$. If $qx = 0$ for some $x \neq 0$, then $R_q \supset S$. So if q is relatively prime to p, since $mp + kq = 1$ for some integers m, k, we would obtain that $s = mps + kqs = 0$, for any $s \in S$, contrary to hypothesis. So any element in R having finite additive order has order a power of p. But since $(p^n - p)x \in C$, we obtain that $(p^{n-1} - 1)p(xy - yx) = 0$; thus $p(xy - yx) = 0$ for all $x, y \in R$.

COROLLARY. *For all $x \in R$, px annihilates all commutators.*

From the first corollary to Theorem 5 we have that $(xy - yx) \in C$. Thus $xy - yx = c$ where $c \in C$ and where $pc = 0$ by Theorem 7. Multiplying from the left by x we obtain $x^2y = xyx + xc = yx^2 + 2xc$. Continuing in this way we obtain that $x^iy = yx^i + icx^{i-1}$. Putting $i = p$, since $pc = 0$ we see that $x^py = yx^p$. Thus we have shown

THEOREM 8. *For every x in R, x^p is in C.*

Since $S \subset N$ we have that $S \subset C$. Since $S \subset$ commutator ideal, which is nilpotent (as can be seen by examining the proof of Theorem 5), $S^2 = (0)$. Let $A(S) = \{x \in R \mid xS = (0)\}$. Clearly $A(S)$ is a two-sided ideal of R, and since it contains S, it is not the zero ideal. Two cases now present themself: (1) $A(S)=R$ and (2) $A(S)\neq R$.

Let us first consider the case that $A(S)=R$. By Theorem 8, for any $x\in R$, $x^p\in C$, so by the corollary to Lemma 3 we have $(x^{np}-x^p)(yz-zy)=0$ for all $x,y\in R$. Thus

$$x^{np-p}x^p(yz-zy)=x^p(yz-zy).$$

Let $T=\{r\in R\,|\,x^{np-p}r=r\}$. Since $x^{np-p}\in C$, T is a two-sided ideal of R. T does not contain S, for $x^{np-p}S=(0)$. Thus $T=(0)$. Since $x^p(yz-zy)\in T$, we get that $x^p(yz-zy)=0$ for all $y,z\in R$. In particular $x^p(xy-yx)=0$. Thus $x^{p+1}y=x^pyx=yx^{p+1}$ since $x^p\in C$. Thus $x^{p+1}\in C$. Continuing in this way we obtain that $x^{p+k}\in C$ for any non-negative k. Since $x^n-x\in C$, $x^{n^p}-x\in C$. But since $n^p>p$, $x^{n^p}\in C$, so $x\in C$. Thus R is commutative; so

THEOREM 9. *If $A(S)=R$, then R is commutative.*

Let us now consider the situation when $A(S)\neq R$. Let $s\in S$, $s\neq 0$. Thus $sR=Rs$ (since $s\in S\subset C$) is a two-sided ideal of R and is contained in S. If $Rs=(0)$, then it would follow that $RS=(0)$ since S is a minimal two-sided ideal, and we have supposed that $RS\neq(0)$. Hence $Rs=S$. Suppose that $x\notin A(S)$. Thus $xs\neq 0$ for $s\neq 0$ in S; for if $xs=0$, then $xsR=xS=0$, a contradiction. Suppose $xs=t\neq 0$. Thus $Rt=S$, whence for some $y\in R$, $s=yt$. Thus $xyt=t$; in fact $xys=s$ for all $s\in S$ by the fact that $tR=S$. So $xy=e$ acts as a unit element on S. For any $x\notin A(S)$ we now can show that we can find a y so that $xy-e\in A(S)$ by the replica of the argument used in finding e. Thus $R/A(S)$ is a field. Hence

THEOREM 10. *$R/A(S)$ is a field.*

Our next goal is to establish that $R/A(S)$ is a finite field. To prove this we must first establish some side results. The first of these is

THEOREM 11. *If $a\in R$ is a divisor of zero then $a\in A(S)$.*

Proof. Suppose that $ax=0$, $x\neq 0$. We maintain that we could have chosen $x\in C$. For if $x^p=0$, then by Lemma 1, $x\in C$, and if $x^p\neq 0$, then by Theorem 8 $x^p\in C$ and $ax^p=0$. So we suppose that $x\in C$. Thus Rx is a two-sided ideal of R. $Rx\neq(0)$ for if $Rx=(0)$, then $T=\{x\in R\,|\,Rx=0\}$ would be a two-sided ideal of R, and not (0), so would contain S, and $RS=(0)$ contrary to hypothesis. So $Rx\neq(0)$, thus $Rx\supset S$, thus $aRx=axR=(0)\supset aS$, and $a\in A(S)$.

This enables us to prove that $R/A(S)$ is a finite field. For if $x\notin A(S)$, since $x^p \varepsilon C$ by Theorem 8, we have that $(x^{np} - x^p)(yz - zy) = 0$ for all $y, z \varepsilon R$. Thus $x^{np} - x^p$ is a divisor of zero and must be in $A(S)$ by Theorem 11. In $R/A(S)$ we then have for every $\bar{x}$, $\bar{x}^{np} = \bar{x}^p$. Thus $R/A(S)$ is a finite field. By the corollary to Theorem 7, for all $x \varepsilon R$, $px \varepsilon A(S)$. So $R/A(S)$ is of characteristic p, and in it $\bar{x}^n = \bar{x}$. This gives

Theorem 12. *$R/A(S)$ is a finite field of characteristic p.*

Corollary. $n \geq p$.

For in $R/A(S)$, $x^n = x$ for all x, and $n > 1$.

Using the fact that the multiplicative group of a finite field is cyclic we immediately obtain

Theorem 13. *There exists an $a \varepsilon R$ so that for every $x \not\varepsilon A(S)$ we can find an integer t so that $x - a^t \varepsilon A(S)$.*

If we could prove that $A(S)$ is in the center, this together with Theorem 13 would lead us quickly to the result we want; consequently the direction we take is to establish that $A(S) \subset C$. Our first result in this regard is

Theorem 14. *If $x \varepsilon A(S)$, then x^p annihilates all commutators.*

Proof. Let $x \varepsilon A(S)$. Since $x^p \varepsilon C$, for all $y, z \varepsilon R$ we get, using Lemma 3, that $(x^{np} - x^p)(yz - zy) = 0$. Thus we have $x^{np-p}x^p(yz - zy) = x^p(yz - zy)$. Let $T = \{r \varepsilon R \mid x^{np-p}r = r\}$. Since $x^{np-p} \varepsilon C$, T is clearly a two-sided ideal of R. S is not in T since $x^{np-p}S = (0) \neq S$. So $T = (0)$, whence $x^p(yz - zy) = 0$, being an element of T. This proves the theorem.

Let $x \varepsilon A(S)$. Thus $x^p(xy - yx) = 0$. So $x^{p+1}y = x^p yx = yx^{p+1}$, since x^p is in the center. Thus $x^{p+1} \varepsilon C$. Continuing in this way we obtain that $x^{p+k} \varepsilon C$ for any non-negative integer k. Since $n \geq p$, we have that $x^n \varepsilon C$, which, combined with $x^n - x \varepsilon C$ yields that $x \varepsilon C$. So we have

Theorem 15. $A(S) \subset C$.

Suppose that $x \not\varepsilon A(S)$. By Theorem 13, $x - a^t \varepsilon A(S) \subset C$; so $x - a^t$ commutes with a hence x commutes with a. Since a also commutes with all the elements in $A(S)$, we see that a is in C. As a consequence of this, $a^t \varepsilon C$, which together with $x - a^t \varepsilon C$ puts x in C. So we have established

Theorem 16. *If $A(S) \neq R$, then R is commutative.*

Theorems 9 and 16 cover between them the only two possibilities when R is subdirectly irreducible. So they yield

THEOREM 17. *If R is subdirectly irreducible and, for every x in R, $x^n - x$ is in the center of R, then R is commutative.*

Using Theorem 6 we obtain the generalization of Jacobson's theorem which we mentioned in the introduction, namely

THEOREM 18. *If R is a ring with center C, and if $x^n - x \varepsilon C$ for all x in R, n a fixed integer larger than 1, then R is commutative.*

Remarks. Before closing we should like to point out two other possible generalizations of Jacobson's theorem which should possibly be true: (1) If $x^{n(x)} - x \varepsilon C$ for all x in R, then R is commutative. (2) If R is a metric ring with a metric ρ and if for every x in R and any $\epsilon > 0$ we can find an integer $n(x, \epsilon) > 1$ so that $\rho(x^{n(x,\epsilon)}, x) < \epsilon$, then R is commutative.

Our original proof used the existence of a unit element. The referee kindly pointed out to us that to prove the theorem without the unit element substantially only involved the case division $A(S) \neq R$ and $A(S) = R$, the latter case not being possible, of course, in rings with unit elements.

THE OHIO STATE UNIVERSITY.

BIBLIOGRAPHY.

[1] G. Birkhoff, "Subdirect unions in universal algebra," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 764-768.

[2] N. Jacobson, "Structure theory for algebraic algebras of bounded degree," *Annals of Mathematics*, vol. 46 (1945), pp. 695-707.

[3] ———, "The radical and semi-simplicity for arbitrary rings," *American Journal of Mathematics*, vol. 67 (1945), pp. 300-320.

A DEFINITION OF DEGREE FOR CERTAIN MAPPINGS IN HILBERT SPACE.*

By Jane Cronin.

1. Introduction. The concept of the degree of a continuous mapping was first introduced by Brouwer [2, 1].[1] Brouwer defined degree for continuous mappings in Euclidean space. A definition of degree for certain mappings in non-locally compact spaces was introduced by Leray and Schauder [8]. They defined a degree for mappings in Banach space of the form $I + F$ where I is the identity transformation and F is completely continuous. Their work left open the question of whether a degree could be defined for all continuous mappings in a Banach space. Leray [7] and Kakutani [6] showed that the answer to this question is no, even if the Banach space is a Hilbert space.

The purpose of this paper is to define a degree for mappings in Hilbert space of the form $I + C + T$, where I is the identity, C is linear, self-adjoint, and completely continuous, and T is a higher order transformation which need not be completely continuous.

In Section 2, the definition of the degree is given and it is proved that this degree has the usual properties of a mapping degree. It is shown that this degree has the same value as the Leray-Schauder degree of $I + C + T$ if the Leray-Schauder degree (hereafter to be termed the LS degree; the Leray-Schauder index [8, p. 54] will be termed the LS index) is defined for $I + C + T$.[2] Also a method for computing the value of the degree for a given mapping is described. The degree defined is essentially the same as the "multiplicity" defined in a previous paper [3]. The real content of Section 2 is the proof that the multiplicity defined in [3] has the properties of a mapping degree, especially that the degree is invariant under homotopy.

The degree is defined only relative to "sufficiently small" sets [3] and in proving the invariance under homotopy of the degree, we assume that the di-

---

* Received May 17, 1950.

[1] Numbers in brackets refer to the bibliography at the end.

[2] In general, the LS degree is not defined for the mapping $I + C + T$ because T is not, in general, completely continuous.

[3] When we refer to the degree at a point p of a mapping M relative to a set S, we mean the degree at p of M/S, i.e., the mapping M considered on the set S.

mension of the null space of $I+C$ is constant during the homotopy. In Section 3, we exhibit an example which justifies the introduction of the degree. That is, we describe a transformation of $L_2(0,1)$ into itself which is of the form $I+C+T$ and hence can be treated by our theory but which is such that its higher order part is not completely continous; so that it is impossible to treat the mapping using the Leray-Schauder theory.

The multiplicity introduced in [3] is actually defined for mappings in Banach space. We are obliged to restrict ourselves to mappings in Hilbert space here in order to prove the invariance under homotopy. Precisely where the Hilbert space theory is used is indicated in footnote [6].

2. Definition of the mapping degree. Let $\mathfrak{X}$ be a Hilbert space. Consider the mapping of $\mathfrak{X}$ into itself $I+C+T$, where I is the identity transformation on $\mathfrak{X}$, C is a linear, self-adjoint completely continuous mapping of $\mathfrak{X}$ into itself and T is a mapping of a subset of $\mathfrak{X}$ into $\mathfrak{X}$ with the following properties:

(P_1) $T(0)=0$;

(P_2) there exist a neighborhood N of 0 and a positive constant B such that if $x_1, x_2 \varepsilon N$, then $\| T(x_1) - T(x_2) \| \leq B[\| x_1 \| + \| x_2 \|][\| x_1 - x_2 \|]$.

Suppose first that $I+C$ is nonsingular, i. e., takes only 0 into 0. Then from the Riesz theory of operators of the form $I+C$ [10], it follows that $(I+C)^{-1}$ exists. Hence the LS index of $I+C$ at 0 is defined and is 1 or -1 (see [8, pp. 56-58]). Now we consider the equation $(I+C+T)(x)=(y)$. Applying $(I+C)^{-1}$ to this equation, we obtain

$$x+(I+C)^{-1}T(x)=(I+C)^{-1}(y). \tag{2.1}$$

From properties P_1 and P_2 of T, it follows that there exist ϵ_1 and ϵ_2 such that if $\| x \| < \epsilon_1$ and $\| y \| < \epsilon_2$, then the Hildebrandt-Graves implicit function theorem [5, Theorem 2] can be applied to solve (2.1) uniquely for x in terms of y. This suggests that we define the mapping degree of $I+C+T$ at y such that $\| y \| < \epsilon_2$ and relative to the sphere $S=[x/\| x \| < \epsilon_1]$[3] to be 1. Actually, to avoid the ambiguity of sign that results from the multiplication by $(I+C)^{-1}$, we define the mapping degree of $I+C+T$ at y such that $\| y \| < \epsilon_2$ and relative to S to be the LS index of $I+C$ at 0. (Pairs $\epsilon'_1 > 0$, $\epsilon'_2 > 0$ such that $\epsilon'_2 \leq \epsilon_1$ and $\epsilon'_2 \leq \epsilon_2$ may be used in place of ϵ_1 and ϵ_2).

Suppose now that $I+C$ is singular, i. e., that there exists $z \varepsilon \mathfrak{X}$, $z \neq 0$ such that $(I+C)(z)=0$. In [10], it is shown that $\mathfrak{X}_1$, the null space of $I+C$, is a finite dimensional linear space. Also, we may write $\mathfrak{X}=\mathfrak{X}_1 \oplus \mathfrak{X}^1$

(direct sum), where $\mathfrak{X}_1$ and $\mathfrak{X}^1$ are orthogonal and on $\mathfrak{X}^1$, $I+C$ is a 1-1 transformation taking $\mathfrak{X}^1$ onto itself. These properties follow from [10] and the fact that C is self-adjoint. If E_1, E^1 are the projections of $\mathfrak{X}$ onto $\mathfrak{X}_1$ and $\mathfrak{X}^1$ respectively, then it is easy to show that the transformation $R=(I+C+E_1)^{-1}$ has the property: $R(I+C)x=E^1(x)$. Moreover, R is of the form $I+G$ where I is the identity and G is linear and completely continuous. Since R is nonsingular and of this form, the LS index at 0 of R is defined.

Now we consider the equation

$$(I+C+T)(x)=y; \tag{*}$$

multiplying by R, we obtain:

$$E^1(x)+RT(x)=R(y). \tag{2. 2}$$

Now we apply E_1 and E^1 respectively to (2. 2) and obtain

$$E_1RT(x_1+x^1)=E_1R(y), \tag{2. 3}$$

where x_1 and x^1 denote $E_1(x)$ and $E^1(x)$ respectively, and

$$x^1+E^1RT(x_1+x^1)=E^1R(y). \tag{2. 4}$$

It is clear that solving (2. 2) (and therefore (*)) for x when y is given is equivalent to solving (2. 3) and (2. 4) simultaneously for x_1 and x^1 when y is given.

From properties P_1 and P_2 of T, it follows that there exist $\epsilon_3>0$, $\epsilon_4>0$, $\epsilon_5>0$ such that if $\| x^1 \| < \epsilon_3$, $\| x_1 \| < \epsilon_4$, $\| y \| < \epsilon_5$, then the Hildebrandt-Graves implicit function theorem can be applied to solve (2. 4) uniquely for x^1 in terms of x_1 and y. That is, we have:

$$x^1=H(x_1,y), \tag{2. 5}$$

where H is uniformly continuous in the subspace of the product space $\mathfrak{X}_1 \times \mathfrak{X}$ in which it is defined. Substituting from (2. 5) into (2. 3), we obtain:

$$E_1RT[x_1+H(x_1,y)]-E_1R(y)=0. \tag{2. 6}$$

Thus the problem of solving (2. 2) for x when y is given is reduced to that of solving (2. 6), an equation in the finite-dimensional space $\mathfrak{X}_1$, for x_1 when y is given. This suggests defining the degree of the mapping $I+C+T$ at a point $y=y_0$ such that $\| y_0 \| < \epsilon_5$ and relative to the sphere $[x/\| x \| < \epsilon_6]$ where ϵ_6 is such that if $\| x \| \leq \epsilon_6$, then $\| x^1 \| < \epsilon_3$ and $\| x_1 \| < \epsilon_4$, to be the topological degree at 0 of the mapping in $\mathfrak{X}_1$:

$$E_1RT[x_1+H(x_1,y_0)]-E_1R(y_0). \tag{2. 7}$$

Actually, in order to avoid the ambiguity of sign resulting from multiplying the original equation by R, we define the degree to be the product of the degree at 0 of (2. 7) and the LS index at 0 of R.

Since the Hildebrandt-Graves theorem yields only a local solution, we must restrict our considerations to small values of x and y. In [3], it was shown that there exist ϵ_7 and ϵ_8 such that if $\| x \| < \epsilon_7$ and $\| y \| < \epsilon_8$, then the Hildebrandt-Graves theorem can be applied and the topological degree of (2. 7) computed.[4] It was also shown that the degree is the same for all y such that $\| y \| < \epsilon_8$. We summarize this discussion in a formal definition of the mapping degree. Let $\epsilon \leq \epsilon_7$, $\epsilon' \leq \epsilon_8$. Then we have:

Definition 2.1. If $I + C$ is nonsingular, the mapping degree of $I + C + T$ at y such that $\| y \| < \epsilon'$ and relative to the sphere $S = [x/\|x\| \leq \epsilon]$ is the LS index at 0 of $I + C$. If $I + C$ is singular, the mapping degree of $I + C + T$ at y such that $\| y \| < \epsilon'$ and relative to S is the product of the LS index at 0 of R and the topological degree of mapping (2. 7) at 0, relative to the sphere $[x_1/\| x_1 \| \leq \epsilon_0]$ in $\mathfrak{X}_1$, where ϵ_0 depends on ϵ as described in [3].[5] We will denote this mapping degree by $d[I + C + T, S, y]$.

Now we show that the mapping degree just defined has some of the usual properties of a mapping degree. First since the degree is defined as the product of the LS index of R and the degree of the mapping

$$E_1RT[x_1 + H(x_1, y)] - E_1R(y)$$

at 0 and relative to the sphere $[x_1/\| x_1 \| \leq \epsilon_0]$, then the mapping degree for sufficiently small y exists if for x_1 such that $\| x_1 \| = \epsilon_0$,

$$E_1RT[x_1 + H(x_1, 0)] \neq 0.$$

Hence the condition that the mapping degree exists is different from the usual condition that the degree of the mapping at y relative to S is defined if no point of the boundary of S is taken into y by the mapping. Instead we have:

THEOREM 2. 1. *The mapping degree* $d[I + C + T, S, y]$ *exists if and only if for each* $x_0 \in S$ *such that* $\| E_1(x_0) \| = \epsilon_0$, *we have* $(I + C + T)(x_0) \neq 0$.

---

[4] In computing the topological degree, properties P_3 and P_4 (stated later on just before Lemma 2. 1) are used. We choose ϵ_7 and ϵ_8 small enough so that P_3 and P_4 can be utilized as described in [3].

[5] From a topological viewpoint, it would be better to call this mapping degree the index of $I + C + T$ at 0. However, in the application, we are interested in nearby points as well as 0 itself. Hence it is more convenient to think of a mapping degree than an index.

Proof. Suppose first that $I+C$ is singular. Then if

$$(I+C+T)x_0=0, \tag{2.8}$$

we have, applying R:

$$E^1(x_0)+RT[E_1(x_0)+E^1(x_0)]=0.$$

Applying E^1 and E_1, we obtain

$$E^1(x_0)+E^1RT[E_1(x_0)+E^1(x_0)]=0 \tag{2.9}$$

and

$$E_1RT[E_1(x_0)+E^1(x_0)]=0. \tag{2.10}$$

Solving (2.9), we have $E^1(x_0)=H[E_1(x_0),0]$. Substituting in (2.10), we have

$$E_1RT\{E_1(x_0)+H[E_1(x_0),0]\}=0. \tag{2.11}$$

Hence the mapping degree is not defined. Reversing the steps in this proof shows that equation (2.11) implies equation (2.8). If $I+C$ is nonsingular, the condition is trivially fulfilled because E_1 is the zero transformation in that case. From the definition, it is known that if $I+C$ is nonsingular, the degree is always defined. This completes the proof of Theorem 2.1.

THEOREM 2.2. *The mapping degree* $d[I+C+T,S,y]$ *has the following properties:*

(i) *If* $d[I+C+T,S,y]\neq 0$, *there exists* $x\in S$ *such that* $(I+C+T)(x)=y$.

(ii) $d[I+C+T,X,y]$ *is constant for all* y *sufficiently close to zero.*

Proof. Property (i) follows directly from the definition of $d[I+C+T,S,y]$ and the corresponding property of the topological degree of the mapping in Euclidean space. If $I+C$ is nonsingular, property (ii) follows from the definition of $d[I+C+T,S,y]$. As remarked before, the proof for the case when $I+C$ is singular is contained in [3, Section 3]. This completes the proof of Theorem 2.2.

THEOREM 2.3. *If* T *is completely continuous and if* $d[I+C+T,S,0]$ *exists, then the* LS *degree of* $I+C+T$ *at* 0 *and relative to* S *exists and is equal to* $d[I+C+T,S,0]$.

Proof. See [3, Section 5].

Now we prove the invariance under homotopy of $d[I+C+T,S,y]$. We state first what we mean by a homotopy in this work.

Definition 2.2. A homotopy $I(x,t)+C(x,t)+T(x,t)$ is a continuous mapping from $\mathfrak{X}\times\mathfrak{I}$, where $\mathfrak{I}$ is the closed unit interval $(0,1)$ into $\mathfrak{X}$ satisfying these conditions:

(1) $I(x,t)=x$ for each pair $(x,t)\varepsilon\,\mathfrak{X}\times\mathfrak{I}$.

(2) For each fixed t, $C(x,t)$ is a linear, self-adjoint, completely continuous transformation of $\mathfrak{X}$ into itself. Also $C(x,t)$ is a continuous function of t in the following sense: for arbitrary $t_0\,\varepsilon\,\mathfrak{I}$, we have $\lim_{t_1\to t_0}\|C_{t_1}-C_{t_0}\|=0$, where C_{t_1} denotes the transformation of $\mathfrak{X}$ into $\mathfrak{X}$ obtained by taking $t=t_1$ in $C(x,t)$ and C_{t_0} is defined similarly.

(3) $T(x,t)$ is a continuous mapping from $\mathfrak{X}\times\mathfrak{I}$ into $\mathfrak{X}$ with the property: for each fixed $t_0\,\varepsilon\,\mathfrak{I}$, $T(x,t)$ is a mapping of a subset of $\mathfrak{X}$ into $\mathfrak{X}$ with properties P_1 and P_2. Moreover the neighborhood N and the constant B are independent of t.

THEOREM 2.4. *If $I(x,t)+C(x,t)+T(x,t)$ is a homotopy satisfying the conditions of Definition 2.2 and if the following additional conditions are satisfied:*

(4) *The dimension of the null space of $I(x,t_0)+C(x,t\ )$ is the same for all $t_0\,\varepsilon\,\mathfrak{I}$.*

(5) *For each $t_0\,\varepsilon\,\mathfrak{I}$ and for a sphere S which is independent of T, $d[I(x,t_0)+C(x,t_0)+T(x,t_0),S,y]$ is defined for y such that $\|y\|<\eta$, where η is independent of t.*

Then the mapping degree $d[I(x,t_0)+C(x,t_0)+T(x,t_0)]$ is the same for all $t_0\,\varepsilon\,\mathfrak{I}$.

Proof. If $I+C$ is nonsingular, the proof follows from the fact that the LS degree of $I+C$ is invariant under homotopy [11, Theorem 2]. So we need only consider the case when $I+C$ is singular. From Definition 2.1 it follows that the mapping degree of $I+C+T$ is the same as the mapping degree of

$$x-E_1(x)+E_1RT[x_1+H(x_1,0)] \tag{2.12}$$

except for a factor ± 1. For the mapping degree of (2.12) is equal to the topological degree at 0 of

$$E_1RT[x_1+H(x_1,0)] \tag{2.13}$$

and $d[I+C+T,S,y]$ for y sufficiently close to 0 is equal to the topological degree of (2.13) multiplied by the LS index at 0 of R, which is 1 or -1. Now (2.12) may be written:

$$(2.14)\qquad x - E_1(x) + E_1RT\{E_1(x) + H[E_1(x), 0]\}.$$

Since E_1 is finite-dimensional and hence completely continuous, it is clear that the LS degree of (2.14) is defined, i.e., (2.14) is a mapping of the form $I + F$ where F is completely continuous.

Now let $E_1{}^t$ be the projection of X onto the null space of $I + C_t$ (for convenience we write the homotopy parameter as a subscript, i.e., we denote $C(x, t)$ by $C_t(x)$ and $T(x, t)$ by $T_t(x)$ and let R_t be the linear transformation such that $R_t(I + C_t)x = x - E_1{}^t(x)$. We let $H_t[E_1{}^t(x), y]$ denote the function obtained by applying Hildebrandt-Graves Theorem.) We will show that

$$(2.15)\qquad x - E_1{}^t(x) + E_1{}^tR_tT_t\{E_1{}^t(x) + H_t[E_1{}^t(x), 0]\}$$

and R_t are homotopies in the sense of [11, Definition 4]. Then the LS degrees are the same for all t such that $0 \leq t \leq 1$ [11]. Since for each t, the mapping (2.15) is a layer mapping [11], it LS degree is by definition the topological degree of $E_1{}^tR_t\{E_1{}^t(x) + H_t[E_,{}^t(x), 0]\}$. The proof of Theorem 2.4 then follows from the relation of the mapping degrees of $I + C + T$ and (2.12) described above.

Hence to complete the proof of Theorem 2.4, it suffices to prove that $E_1{}^t$, R_t, and H_t are continuous mappings from $\mathfrak{X} \times \mathfrak{I}$ into $\mathfrak{X}$ and that the mapping

$$E_1{}^t\{-x + R_tT_t\{E_1{}^t(x) + H_t[E_1{}^t(x), 0]\}\}$$

takes any bounded set in $\mathfrak{X} \times \mathfrak{I}$ into a sequentially compact set in $\mathfrak{X}$. (Cf. [11, Definition 4].) This last statement is an immediate consequence of the fact that any bounded set $V = (y_\nu, t_\nu)$ of $\mathfrak{X} \times \mathfrak{I}$ each of whose points satisfy the equation $-y_\nu = C_{t_\nu}(y_\nu)$ is sequentially compact in $\mathfrak{X} \times \mathfrak{I}$. To prove this: let t_0 be a limit point of the set $\{t_\nu\}$. Then there is a subsequence y_{ν_n} such that $\lim\limits_{n\to\infty} C_{t_0}(y_{\nu_n})$ exists and is some point, say y_0. This is because C_{t_0} is completely continuous and $\{y_\nu\}$ is bounded. Then by (2) of Definition 2.2, there exists $\lim\limits_{n\to\infty} C_{t_{\nu_n}}(y_{\nu_n}) = -\lim\limits_{n\to\infty} y_{\nu_n} = y_0$.

Now we need to show that $E_1{}^t$ is a continuous mapping from $\mathfrak{X} \times \mathfrak{I}$ into $\mathfrak{X}$.[6] More precisely, we want to show $\lim\limits_{t\to t_0} \| E_1{}^t - E_1{}^{t_0} \| = 0$. This is a direct consequence of [9, Satz 5]. By definition, $R_t = (I + C_t + E_1^t)^{-1}$; so R_t is a continuous mapping from $\mathfrak{X} \times \mathfrak{Y}$ into $\mathfrak{X}$.

---

[6] It is for the proof of this statement that Hilbert space theory is used here.

Finally, we must show that H_t depends continuously on t. $H_t[E_1^t(x), y]$ is the solution of the equation

$$(2.17) \qquad E_t^1(x) + E_t^1 R_t T_t[E_1^t(x) + E_t^1(x)] = E_t^1 R_t(y)$$

for $E_t^1(x)$ in terms of y and $E_1^t(x)$ which is obtained by using the implicit function theorem of Hildebrandt and Graves [5, Theorem 2]. For fixed t_0, the initial solution of (2.17) is $E_t^1(x) = E_1^t(x) = y = 0$, and $H_t[E_1^t(x), y]$ is the limit of the sequence:

$$H_{t_0}^{(1)}[E_1^{t_0}(x), y] = E_{t_0}^1 R_{t_0} T_{t_0}[E_1^{t_0}(x)] - E_{t_0}^1 R_{t_0}(y)$$
$$H_{t_0}^{(2)}[E_1^{t_0}(x), y] = E_{t_0}^1 R_{t_0} T_{t_0}\{E_1^{t_0}(x) + H_{t_0}^{(1)}[E_1^{t_0}(x), y]\} - E_{t_0}^1 R_{t_0}(y)$$
$$\cdot \quad \cdot \quad \cdot$$
$$H_{t_0}^{(n)}[E_1^{t_0}(x), y] = E_{t_0}^1 R_{t_0} T_{t_0}\{E_1^{t_0}(x) + H_{t_0}^{(n-1)}[E_1^{t_0}(x), y]\} - E_{t_0}^1 R_{t_0}(y).$$

In [5], it is proved that the sequence $H_{t_0}^{(n)}[E_1^{t_0}(x), y]$ is a Cauchy sequence for each fixed pair $[E_1^{t_0}(x), y]$ and that the convergence is uniform in $[E_1^{t_0}(x), y]$. To prove the desired statement, we show that the convergence is uniform in t. Since $\| E_1^t \|$, $\| E_t^1 \|$, and $\| R_t \|$ are bounded in t, and since by (3) of Definition 2.2, the bound B of T_t is independent of t, the same type of argument as is used in [5] can be used to show that the convergence is uniform in t. This completes the proof of Theorem 2.4.

Now we describe a method for computing the degree of a given mapping. In order to do this, we impose two more conditions on the transformation. We assume the transformation T has the following properties:

(P_3) The transformation T is split into a term of order $k (\geq 2)$ and a term of higher order; that is, $T(x) = T^{(k)}(x) + \tilde{T}^{(k+1)}(x)$ where $T^{(k)}$ and $\tilde{T}^{(k+1)}$ are both continuous and:

(1) $T^{(k)}$ is homogeneous of degree k; that is, if m is an arbitrary integer, then

$$T^{(k)}(\sum_{i=1}^{m} a_i x_i) = \sum_{(b_i)} a_1^{b_1} \cdots a_m^{b_m} T b_1 \cdots b_m [x_1, \cdots, x_m],$$

where the summation is over all sets of non-negative integers $[b_1, \cdots, b_m]$ such that $\sum_{i=1}^{m} b_i = k$ and $T_{b_1 \ldots b_m}$ is a continuous mapping from $\mathfrak{X} \times \cdots \times \mathfrak{X}$ (m times) into $\mathfrak{X}$.

(2) The $\lim_{x\to 0} \| \tilde{T}^{(k+1)}(x) \| / \| x \|^k = 0$.

(P_4) $E_1 R T^{(k)}(x_1) \neq 0$ for all $x_1 \varepsilon \mathfrak{X}_1$ such that $x_1 \neq 0$.

LEMMA 2.1. *If $I + C$ is singular and if T has properties P_1, P_2, P_3, and P_4, then the mapping degree of $I + C + T$ is equal to the topological degree at zero of a mapping of Euclidean n-space E^n (real or complex as $\mathfrak{X}$ is real or complex) given by n polynomials in n variables, the polynomials homogeneous of degree k in the n variables, where n is the dimension of the null space of $I + C$ and k is the "order" of T described by P_3.*

Proof. See [3, Section 3].

The degree of such a mapping described by homogeneous polynomials can be computed. We have:

THEOREM 2.5. *The topological degree at zero of the type of polynomial mapping described in Lemma 2.1 is k^n if E^n is complex and is $\equiv k^{n-1} \pmod 2$ if E^n is real.*

The proof of Theorem 2.5 will be given in another paper.

3. An example. In order to justify the definition of mapping degree given in Section 2, we must show that it is independent of previous work. That is, we must show that our theory is not contained in the Leray-Schauder theory. So we must exhibit a mapping $I + C + T$ which satisfies our hypotheses but which is not of the form $I + F$ where F is completely continuous. Since C is completely continuous, we must give an example of a mapping T which has properties P_1 and P_2 and which is not completely continuous.

A simple such example is the following: Let $\mathfrak{X}$ be the Hilbert space of square Lebesgue integrable functions $x(t)$ over the group of real numbers mod 1, i. e., $\mathfrak{X} = L_2(0,1)$. We define

$$T[x(t)] = \int_0^1 x(t)x(s-t)\,dt$$

i. e., $T[x(t)]$ is the convolution of x with itself. It is a straightforward verification to show that this T has properties P_1 and P_2. The transformation T is not completely continuous because the bounded set $\{e^{2\pi int}\}$ is taken into itself and it is clear that $\{e^{2\pi int}\}$ contains no convergent subsequence.

Examples of the transformation C abound. Any linear integral transformation acting on $x(t) \varepsilon L_2(0,1)$,

$$\int_0^1 K(x,t)x(t)\,dt,$$

where $K(s,t) = \overline{K(t,s)}$ and $K(s,t)$ is continuous simultaneously in s and t, is self-adjoint and completely continuous. Hence as a simple example of a transformation to which our theory may be applied but to which the Leray-Schauder theory cannot be applied, we have the transformation

$$x(s) + \int_0^1 K(s,t)x(t)\,dt + \int_0^1 x(t)x(s-t)\,dt$$

acting on $x(s) \,\varepsilon\, L_2(0,1)$.

ANN ARBOR, MICHIGAN.

BIBLIOGRAPHY.

[1] P. Alexandroff and H. Hopf, *Topologie* I, Berlin, 1935.

[2] L. E. J. Brouwer, "Über Abbildung von Mannigfaltigkeiten," *Mathematische Annalen*, vol. 71 (1911), pp. 97-115.

[3] J. Cronin, "Branch points of solutions of equations in Banach space," *Transactions of the American Mathematical Society*, vol. 69 (1950), pp. 208-231.

[4] P. Halmos, *Finite Dimensional Vector Spaces*, Princeton University Press, 1942.

[5] T. H. Hildebrandt and L. M. Graves, "Implicit functions and their differentials in general analysis," *Transactions of the American Mathematical Society*, vol. 29 (1927), pp. 127-153.

[6] S. Kakutani, "Topological properties of the unit sphere of a Hilbert space," *Proceedings of the Imperial Academy of Tokyo*, vol. 19 (1943), pp. 269-271.

[7] J. Leray, "Les problèmes non-linéaires," *L'enseignement mathématique*, vol. 35 (1936), pp. 139-151.

[8] J. Leray and J. Schauder, "Topologie et équations fonctionnelles," *Annales de l'École Normale*, vol. 51 (1934), pp. 45-78.

[9] F. Rellich, "Störungstheorie der Spektralzerlegung II," *Mathematische Annalen*, vol. 113 (1937), pp. 677-685.

[10] F. Riesz, "Über lineare Funktionalgleichungen," *Acta Mathematica*, vol. 41 (1918), pp. 71-98.

[11] E. H. Rothe, "The theory of topological order in some linear topological spaces," *Iowa State College Journal of Science*, vol. 13 (1939), pp. 373-390.

ON APPROXIMATION BY NON-DENSE SETS OF TRANSLATES.*

By I. I. HIRSCHMAN, JR.

We denote by $L_2(-\infty,\infty)$ the Hilbert space of complex valued functions $f(x)$ defined for $-\infty < x < \infty$ and such that

$$\| f(x) \| = [\int_{-\infty}^{\infty} | f(x) |^2 dx]^{1/2}$$

is finite. It has been shown by Laurent Schwartz [4], that if $t_1 < t_2 < t_3 < \cdots$ is a sequence of real numbers such that

$$\sum_{n=1}^{\infty} e^{-t_n} < \infty, \liminf_{n\to\infty} (t_{n+1} - t_n) e^{t_n} > 0,$$

and if $f(x)$ lies in the closure in $L_2(-\infty,\infty)$ of the subspace of finite linear combinations of the functions

$$e^{\frac{1}{2}(t_n-x)}\exp[-e^{(t_n-x)}], \qquad (n = 1, 2, \cdots),$$

i. e., if given $\epsilon > 0$ there is an integer N and constants $a_1, a_2, \cdots, a_N$ such that

$$\| f(x) - \sum_{n=1}^{N} a_n e^{\frac{1}{2}(t_n-x)}\exp[-e^{(t_n-x)}] \| \leqq \epsilon,$$

then $f(x)$ can be represented as a series of the form

$$f(x) = \sum_{n=1}^{\infty} b_n e^{\frac{1}{2}(t_n-x)}\exp[-e^{(t_n-x)}]$$

for almost all x. In the present paper we shall establish a general theorem of this type.

Let $\Phi(x) \varepsilon L_2(-\infty,\infty)$ and $\phi(t) \varepsilon L_2(-\infty,\infty)$ be Fourier transforms of each other.

$$\Phi(x) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{itx}\phi(t)\,dt \qquad (\mathrm{M}_2),$$

(1)

$$\phi(t) = (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{-itx}\Phi(x)\,dx \qquad (\mathrm{M}_2).$$

---

* Received September 16, 1950. Research supported by the Office of Naval Research.

The symbol M_2 indicates that the integral in question is to be taken as the limit in the mean of order two of the partial integrals $\int_{-T}^{T}$ as T increases to infinity. Let $\nu(t)$ be defined for $0 \leqq t < \infty$ and such that

$$\nu(0) > 0, \qquad \nu'(t) \geqq 0, \qquad (0 \leqq t < \infty),$$
$$(2) \qquad t\nu'(t)/\nu(t) = o(1) \text{ as } t \to +\infty, \int_1^{\infty} [t\nu(t)]^{-1} dt = \infty.$$

We define $H(t)$ and $\mathfrak{H}(t)$ by the equations

$$H(t) = (2/\pi) \int_1^t [u\nu(u)]^{-1} du,$$
$$(3) \qquad t = (2/\pi) \int_1^{\mathfrak{H}(t)} [u\nu(u)]^{-1} du.$$

It is supposed that $\{t_n\}_1^{\infty}$ is an increasing sequence of real numbers.

THEOREM. *Let $\Phi(x)$, $\phi(t)$, $\nu(t)$, $H(t)$, $\mathfrak{H}(t)$, and $\{t_n\}_1^{\infty}$ be defined as above. If*

1. $|\phi(t)| \geqq \exp[-|t|/\nu(|t|)]$ *for all t,*
2. $|\Phi(x)| \leqq \exp[-\mathfrak{H}(x)]$ *for large positive x,*
3. $\liminf_{n\to\infty} (t_{n+1} - t_n) n\nu(n) > 0,$
4. $\liminf_{n\to\infty} t_n/H(n) > 1,$

and if $f(x)$ lies in the closure of the subspace of $L_2(-\infty, \infty)$ formed by the finite linear combinations of the functions $\{\Phi(t_n - x)\}_1^{\infty}$, then $f(x)$ may be represented as a series of the form

$$f(x) = \sum_{n=1}^{\infty} b_n \Phi_n(t_n - x)$$

for almost all x.

Our proof depends upon the construction of a sequence of functions $\{\Psi_n(x)\}_1^{\infty}$ such that

$$(4) \qquad (2\pi)^{-1/2} \int_{-\infty}^{\infty} \Psi_n(x) \Phi(t_m - x)\, dx = \delta_{n.m},$$

$$(5) \qquad \| \Psi_n \| = O(\exp \mathfrak{H}(\mu t_n)), \qquad n \to \infty, \mu < 1.$$

Suppose that we have exhibited a sequence of continuous functions $\Lambda_n(t)$, $n=1,2,\cdots$, defined for $-\infty < t < \infty$ and belonging to $L_2(-\infty,\infty)$, such that if $\lambda_n(u) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} e^{itu}\Lambda_n(t)\,dt$ then

$$\Lambda_n(t_m) = \delta_{n,m}, \tag{6}$$

$$\| \lambda_n(u)/\phi(u) \| = O(\exp \mathfrak{H}(\mu t_n)), \qquad n \to \infty,\ \mu < 1. \tag{7}$$

We set $\Psi_n(t) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} e^{-itu}[\lambda_n(u)/\phi(u)]\,du$ (M_2). It is asserted that the functions $\{\Psi_n\}_1^\infty$ so defined satisfy conditions (4) and (5). By Parseval's equality we have

$$\int_{-\infty}^{\infty} \Psi_n(x)\Phi(t_m - x)\,dx = \int_{-\infty}^{\infty} [\lambda_n(u)/\phi(u)][\phi(u)e^{-it_m u}]\,du$$

$$= \int_{-\infty}^{\infty} \lambda_n(u)e^{-iut_m}\,du = (2\pi)^{1/2}\Lambda_n(t_m) = (2\pi)^{\frac{1}{2}}\delta_{n,m}.$$

Using Parseval's equality again we obtain

$$\| \Psi_n(x) \| = \| \lambda_n(u)/\phi(u) \| = O(\exp \mathfrak{H}(t_n\mu)), \qquad n \to \infty.$$

We proceed to the construction of the functions $\{\Lambda_n(t)\}_1^\infty$. By assumption 4 there exists a number r, $0 < r < 1$, such that for n sufficiently large $t_n \geqq H(n)/r^2$. We define $\psi(t)$, $\mathfrak{H}^*(t)$, $R(t)$ and $\rho(t)$ by the equations:

$$\begin{aligned} &\psi(t) = v(t)r^2, \qquad t = (2/\pi)\int_1^{\mathfrak{H}^*(t)} [u\psi(u)]^{-1}du, \\ &R(t) = \log \mathfrak{H}^*(t), \qquad \rho(t) = \psi[\mathfrak{H}^*(u)]. \end{aligned} \tag{8}$$

Let S_z be the strip in the z plane, $z = x + iy$, defined by the inequality $|y| < \pi/2$ for $-\infty < x < \infty$, and let S_w be the curvilinear strip in the w plane, $w = u + iv$, defined by the inequalities $|v| < r/\rho(0)$ for $-\infty < u < 0$, $|v| < r/\rho(u)$ for $0 \leqq u < \infty$. Let $\chi(w)$ be the function which maps S_w conformally upon S_z and is normalized by the conditions $\chi(0) = 0$, $\chi'(0) > 0$.

We set

$$\begin{aligned} &d = \exp(4\pi r)/\cos(\pi r/2) \\ &m(u) = \exp\{-d\exp[r\chi(w)]\}/[4\rho(0)^2 + u^2]. \end{aligned} \tag{9}$$

Let $A^*(z)$ be analytic and $|A^*(z)| \leqq 1$ in S_z. We define

$$(10)\qquad \begin{aligned} A(u) &= m(u)A^*[\chi(u)] \\ \alpha(t) &= (2\pi)^{-1/2}\int_{-\infty}^{\infty} A(u)e^{itu}du, \qquad -\infty < t < \infty. \end{aligned}$$

It has been shown in [2, § 5] that under these assumptions

$$(11)\qquad |\alpha(t)| \leqq c\exp[-|t|/r\nu(|t|)], \qquad -\infty < t < \infty,$$

where c is a constant depending on $\nu(t)$ and r, but independent of $A^*(z)$.

Using (2) and (8) it is easily verified that

$$(d/dt)[r/\rho(t)] = -(r\pi/2)\mathfrak{H}^*(t)\nu'[\mathfrak{H}^*(t)]/\nu[\mathfrak{H}^*(t)] = o(1),$$

as $t \to +\infty$. It follows from [3; § 2] that

$$(12)\qquad \begin{aligned} &\chi(t) \sim R(t)/r, \qquad t \to +\infty, \\ &\lim_{t\to+\infty} \chi(t+\rho(t)^{-1}wr) - \chi(t) = (\pi/2)w, \end{aligned}$$

uniformly for $|w| \leqq 1-\epsilon$, $\epsilon > 0$.

Let $x_n = \chi(t_n)$, $n = 1, 2, \cdots$. By (12) we have $x_n \sim R(t_n)/r$ as $n \to \infty$. Now $R(t) = \log \mathfrak{H}^*(t)$, $\mathfrak{H}^*(t) = \mathfrak{H}(r^2t)$, $t_nr^2 \geqq H(n)$ for large n, and $\mathfrak{H}(t)$ is increasing. It follows that x_n asymptotically exceeds $\log n/r$. From this we deduce that $\Sigma e^{-x_n} < \infty$. By assumption 4 $\liminf n\nu(n)[t_{n+1} - t_n] > 0$ as $n \to \infty$. We have $\rho(t_n) = r^2\nu[\mathfrak{H}(r^2t_n)]$. Since $t_n \geqq r^{-2}H(n)$ for large n and since $\mathfrak{H}(t)$ is increasing we have $\rho(t_n) \geqq r^2\nu(n)$. Thus as $n \to \infty$ $\liminf n\rho(t_n)[t_{n+1} - t_n] > 0$. It follows from (12) that $\liminf n[x_{n+1} - x_n] > 0$. Using the fact that x_n asymptotically exceeds $\log n/r$, we obtain finally that $\liminf (x_{n+1} - x_n)e^{x_n} > 0$.

We define

$$L^*_n(z) = \prod_{m=1}^{\infty}{}' [e^z - {}^{x_m}]/[e^z + e^{x_m}], \ (m \neq n), \quad n = 1, 2, \cdots,$$

Since $\sum_{n=1}^{\infty} e^{-x_n} < \infty$, we have $L^*(z)$ analytic and $|L^*(z)| \leqq 1$ in S_z. See [4, § 9, and § 10]. We put

$$(13)\qquad L_n(u) = m(u)L^*_n[\chi(u)], \qquad -\infty < u < \infty.$$

We may now define the functions $\{\Lambda_n\}_1^{\infty}$. We set

$$(14)\qquad \Lambda_n(t) = L_n(t)/L_n(t_n), \qquad n = 1, 2, \cdots.$$

We must verify properties (6) and (7). Property (6) is evident. To establish (7) we note that the relation $\liminf_{n\to\infty}[x_{n+1}-x_n]e^{x_n} > 0$ implies that $|L^*_n(x_n)| \geqq \exp[-\exp(x_n - a)]$ for any a, provided n is sufficiently large. See [4, § 9 and § 10]. Taking $a = 0$ we have $L^*_n[\chi(t_n)] \geqq \exp[-\exp x_n]$ for n sufficiently large. Now $x_n \sim R(t_n)/r$ as $n\to\infty$, $R(t_n) = \log \mathfrak{H}^*(t_n)$, $\mathfrak{H}^*(t_n) = \mathfrak{H}(r^1 t_n)$. Thus $\log\log|L^*[\chi(t_n)]|^{-1}$ is asymptotically less than $r^{-1}\log\mathfrak{H}(r^2 t_n)$ as $n\to\infty$. Again $m(t_n)$ asymptotically exceeds

$$t_n^{-2}\exp\{-d\exp[r\chi(t_n)]\},$$

so that $\log\log[1/m(t_n)]$ is asymptotically less than $\log\mathfrak{H}(r^2 t_n)$ as $n\to\infty$. Using the relation $\log\mathfrak{H}(r^2 t_n)$ is asymptotically less than $r^{3/2}\log\mathfrak{H}(r^{1/2}t_n)$ as $n\to\infty$, see [2; § 2], we obtain $|L_n(t_n)| \geqq \exp[-\mathfrak{H}(r^{1/2}t_n)]$ for n sufficiently large. Referring to (11) we see that if $\lambda_n(u) = (2\pi)^{-1/2}\int_{-\infty}^{\infty} e^{itu}\Lambda_n(t)dt$, then

$$|\lambda_n(u)| \leqq c'\exp[-|u|/r\nu(|u|)]\exp[+\mathfrak{H}(r^{1/2}t_n)].$$

Using assumption 1, we obtain (7) with $\mu = r^{1/2}$.

We can now complete our proof. Let $f(x)$ be any function in $L_2(-\infty,\infty)$. We associate with $f(x)$ the formal series $S(f) = \sum_{n=1}^{\infty} b_n(f)\Phi(t_n - x)$, where

$$b_n(f) = (2\pi)^{-1/2}\int_{-\infty}^{\infty}\Psi_n(x)f(x)\,dx.$$

By (5) and Schwarz's inequality we see that for x fixed

$$|b_n(f)\Phi(t_n - x)|/\|f\| = O(\exp[-\mathfrak{H}(t_n - x) + \mathfrak{H}(\mu t_n)]), \qquad n\to\infty.$$

Choose μ', $\max(\mu, r^2) < \mu' < 1$. If n is sufficiently large $t_n - x \geqq \mu' t_n$, and hence $-\mathfrak{H}(t_n - x) + \mathfrak{H}(\mu t_n) \leqq -\mathfrak{H}(\mu' t_n) + \mathfrak{H}(\mu t_n)$. By [2; § 2] we have $\log\mathfrak{H}(\mu t_n)$ is asymptotically less than $(\mu/\mu')\log\mathfrak{H}(\mu' t_n)$, from which it follows that $-\mathfrak{H}(\mu' t_n) + \mathfrak{H}(\mu t_n) \sim -\mathfrak{H}(\mu' t_n)$ as $n\to\infty$. Finally $\mu' t_n \geqq \mathrm{H}(n)$ for n sufficiently large. Combining our estimates we have $-\mathfrak{H}(t_n - x) + \mathfrak{H}(\mu t_n)$ is asymptotically less than $-n$ as $n\to\infty$. It follows that there exists a function $M(x)$ finite almost everywhere depending only on x and $\{t_n\}_1^{\infty}$ and $\nu(t)$, but independent of $f(x)$ such that $|S(f)| \leqq M(x)\|f\|$, $-\infty < x < \infty$. We now assume that $f(x)$ lies in the closure of the linear subspace of $L_2(-\infty,\infty)$ formed by the finite linear combinations of the functions $\{\Phi(t_n - x)\}_1^{\infty}$. There will then exist a sequence $\{F_k(x)\}^{\infty}_{k=1}$ of functions of the form

$$F_k(x) = \sum_{n=1}^{N_k} b_n^{(k)}\Phi(t_n - x)$$

such that $\lim \| f(x) - F_k(x) \| = 0$. It is no restriction to suppose that $\lim F_k(x) = f(x)$ almost everywhere since we can certainly select a subsequence with this property, and, if necessary, replace our original sequence by the subsequence. We have

$$| S(f) - S(F_k) | = | S(f - F_k) | \leqq \| f - F_k \| M(x).$$

But $S(F_k) = F_k$; thus $S(f) = \lim F_k = f(x)$ for almost all x. Q. E. D.

If we set $\nu = 1$, the special case of our theorem which results may be applied to $e^{x/2} \exp[-e^x]$ to give Schwartz's theorem with the conditions

$$\sum_{n=1}^{\infty} e^{-t_n} < \infty, \liminf_{n \to \infty} [t_{n+1} - t_n] e^{t_n} > 0,$$

replaced by the more restrictive conditions

$$\liminf_{n \to \infty} t_n / \log n > 1, \liminf_{t = +\infty} [t_{n+1} - t_n] n > 0.$$

WASHINGTON UNIVERSITY.

BIBLIOGRAPHY.

[1] I. I. Hirschman, Jr., "On the behaviour of Fourier transforms at infinity and on quasi-analytic classes of functions," *American Journal of Mathematics*, vol. 72 (1950), pp. 200-213.

[2] ———, "On the distributions of the zeros of functions belonging to certain quasi-analytic classes," *ibid.*, pp. 396-406.

[3] ———, "On non-uniformly quasi-analytic functions," *ibid.*, pp. 863-867.

[4] L. Schwartz, *Études des sommes d'exponentielles réelles*, Paris, 1943.

TAUBERIAN THEOREMS FOR SUMMABILITY (R_1).*

By Otto Szász.

---

1. Introduction. A series $\sum_0^\infty a_n$ is called summable (R_1) to the value s if the series

$$\sum_1^\infty s_n n^{-1} \sin nh, \qquad s_n = \sum_0^n a_\nu \tag{1.1}$$

converges in some interval $0 < h < h_0$, and if

$$2/\pi \sum_1^\infty s_n n^{-1} \sin nh \to s \text{ as } h \to 0. \tag{1.2}$$

It is convenient to assume $a_0 = 0$.

The method (R_1) is not regular, as the regularity condition $\sum_1^\infty n^{-1} |\sin nt| <$ constant is not satisfied.

If the series (1.1) converges in $0 < h < h_0$, then $s_n n^{-1} \sin nh \to 0$, $0 < h < h_0$, hence, by a well known theorem of G. Cantor,

$$s_n/n \to 0, \qquad n \to \infty. \tag{1.3}$$

If $s_n \equiv 1$, then $\sum_1^\infty n^{-1} \sin nh = (\pi - h)/2 \to \pi/2$, so that (1.2) holds.

A related method is summability $(R, 1)$, sometimes called Lebesgue's method. If the series $\sum_1^\infty a_n n^{-1} \sin nh$ is convergent in some interval $0 < h < h_0$, and if $\sum a_n (nh)^{-1} \sin nh \to s$, $h \to 0$, then $(R, 1) \sum a_n = s$. Several Tauberian theorems can be found in the literature, that is conditions under which summability $(R, 1)$ implies convergence and vice versa. We give here similar theorems for summability (R_1). Recently Hardy and Rogosinski [1] gave necessary and sufficient conditions for a function $f(t)$ in order that its Fourier series be summable (R_1). Our conditions are mainly on the terms a_n of a series $\sum a_n$.

2. Some lemmas.

Lemma 1. *If the series* (1.1) *converges in* $0 < h < h_0$, *then*

$$\sum_1^\infty s_n n^{-1} \sin nh = \sum_1^\infty a_n \rho_n(h), \tag{2.1}$$

---

* Received August 14, 1950. Presented at the International Congress of Mathematicians, Cambridge, Massachusetts. The preparation of this paper was sponsored (in part) by the Office of Naval Research.

where

$$\rho_n(h) = \sum_{\nu=n}^{\infty} \nu^{-1} \sin \nu h. \tag{2. 2}$$

Conversely, if $n^{-1}s_n \to 0$, then the convergence of the right hand side implies (2. 1).

Partial summation yields

$$\sum_{n}^{m} s_\nu - 1 \sin \nu h = s_n\rho_n - s_m\rho_{m+1} + \sum_{n+1}^{m} a_\nu \rho_\nu, \tag{2. 3}$$

and $\rho_n(h) = -(\tau_{n-1})/n + \sum_{n}^{\infty} \tau_\nu/[\nu(\nu+1)]$, where

$$\tau_n(h) = \sum_{1}^{n} \sin \nu h = (\sin nh)/2 + \cot(h/2)\sin^2(nh/2), \qquad 0 < h < \pi.$$

It follows that $|\rho_n(h)| < n^{-1} |\tau_{n-1}| + \sum_{n}^{\infty} [\nu(\nu+1)]^{-1} |\tau_\nu|$. Now $|\tau_n(h)| < 1/2 + [\sin(h/2)]^{-1}$, hence

$$|\rho_n(h)| < n^{-1}[1 + 2/\sin(h/2)], \tag{2. 4}$$

and, in view of (1. 3), $s_n\rho_{n+1} \to 0$, for fixed h, as $n \to \infty$. Thus $\sum a_\nu\rho_\nu$ converges, and

$$\sum_{n}^{\infty} s_\nu \nu^{-1} \sin \nu h = \sum_{n+1}^{\infty} a_\nu \rho_\nu(h) + s_n\rho_n(h)\,; \tag{2. 5}$$

$n = 1$ yields (2. 1). The converse follows similarly from (2. 3) and (2. 4).

We shall call a series $\sum a_n$ summable $(\bar{R}_1)$ to s if $\sum_{1}^{\infty} a_n\rho_n(h)$ converges for small positive h, and if $2/\pi \sum_{1}^{\infty} a_n\rho_n(h) \to s$, $h \to 0$. It follows from Lemma 1 that summability $(\bar{R}_1)$ and $n^{-1}s_n \to 0$ imply summability (R_1).

LEMMA 2. *Suppose that* $0 < \delta \leq 1$,

$$\sum_{n}^{2n} (|a_\nu| - a_\nu) = O(n^{1-\delta}), \tag{2. 6}$$

and

$$s_n = \sum_{0}^{n} a_\nu = O(n^{1-\delta})\,; \tag{2. 7}$$

then $\sum_{1}^{\infty} \nu^{-1} |a_\nu|$ *converges, and* (see [3], p. 579)

$$\sum_{n}^{\infty} \nu^{-1} |a_\nu| = O(n^{-\delta}). \tag{2. 8}$$

LEMMA 3. *If* $\sum a_n$ *is Abel-summable, and if*

$$\sum_{n}^{2n} (|a_\nu| - a_\nu) = O(1), \tag{2. 9}$$

then $\sum a_n$ *is summable* $(C, 1)$ (*Cesàro mean of order* 1), *and*

$$s_n = O(1). \tag{2.10}$$

See [4], Lemmas 1 and 3.

3. Tauberian conditions for summability (R_1).

THEOREM 1. *If* (2.9) *holds, then Abel-summability of* $\sum a_n$ *implies summability* (R_1) *to the same value.*

Let $\sum_1^n s_\nu = s'_n$; by Lemma 3, $n^{-1}s'_n$ tend to a limit s; we may assume without loss of generality that $s = 0$ (otherwise replace a_1 by $a_1 - s$).

We write

$$\sum_1^\infty s_\nu \nu^{-1} \sin \nu h = \sum_1^n + \sum_{n+1}^\infty = A_n(h) + B_n(h), \text{ say.} \tag{3.1}$$

Summation by parts gives

$$A_n(h) = s'_n n^{-1} \sin nh + \sum_1^{n-1} s'_\nu [(\sin \nu h)/\nu - \sin(\nu+1)h/(\nu+1)].$$

But

$$\nu^{-1} \sin \nu h - \sin(\nu+1)h/(\nu+1)$$
$$= \sin(\nu+1)h/\nu(\nu+1) - (2/\nu)\sin(h/2)\cos(2\nu+1)h/2,$$

hence

$$|(\sin \nu h)/\nu - \sin(\nu+1)h/(\nu+1)| < h/\nu + h/\nu = 2h/\nu.$$

Given a positive $\epsilon < 1/2$, choose $n_0(\epsilon)$ so that $|s'_n| < \epsilon^2 n/2$ for $n > n_0$; then

$$\sum_1^n \nu^{-1} |s'_\nu| < \epsilon^2 n, \text{ for } n > n_1(\epsilon) > n_0.$$

To any h in the interval $0 < h < 1/n_1$ choose $n = 1 + (\epsilon h)^{-1}$, so that

$$n > (\epsilon h)^{-1} > n_1/\epsilon > 2n_1, \text{ and } n < 1 + (\epsilon h)^{-1}. \tag{3.2}$$

Now

$$|A_n(h)| < \epsilon^2 + 2h \sum_1^n \nu^{-1} |s'_\nu| < \epsilon^2(1 + 2hn)$$
$$< \epsilon^2(1 + 2h + 2/\epsilon) < 2\epsilon + 3\epsilon^2.$$

Next, using (2.3) and (2.4) we get convergence of the series (3.1), and from (2.4) and (2.5)

$$(3.3)\quad B_n(h) < |s_{n+1}|(n+1)^{-1}[1+2/\sin(h/2)] + [1+2/\sin(h/2)]\sum_{n}^{\infty}|a_\nu|/\nu.$$

From (2.10) and (3.2), $s_{n+1}(n+1)^{-1}[1+2/\sin(h/2)] = O(\epsilon)$, and from (2.8) for $\delta = 1$,

$$(3.4)\qquad \sum_{n}^{\infty}\nu^{-1}|a_\nu| = O(1/n).$$

It follows that $\sum_{1}^{\infty} s_\nu(\sin \nu h)/\nu = O(\epsilon)$ for $h < h_0(\epsilon)$. This proves Theorem 1.

The analogous theorem holds for summability $(R, 1)$; see [4,], Theorem 4. There is also given a Fourier power series, satisfying the assumptions, while $\sum a_n$ diverges.

THEOREM 2. *If* (2.9) *holds and if* $\sum a_n$ *converges, then* (1.1) *converges uniformly in* $0 \leq h \leq \pi$. *In particular* $\sum a_n$ *is summable* (R_1).

We may assume that $s_n \to 0$; we shall show that $B_n(h)$ is uniformly small for large n. We write

$$B_n(h) = \sum_{n+1}^{\lambda n} + \sum_{\nu > \lambda n}, \text{ where } \lambda > 1.$$

From (3.3) and (3.4), $\sum_{\nu > \lambda n} s_\nu \nu^{-1} \sin \nu h = O(\lambda n h)^{-1}$. Given $\epsilon > 0$, choose $n_0(\epsilon)$ so that $|s_n| < \epsilon^2$ for $n > n_0$; define $k = k(\lambda)$ by $n + k = [\lambda n]$, so that $k \leq (\lambda - 1)n$. To any h in $0 < h < \pi$ choose $\lambda = 1 + (\epsilon n h)^{-1}$. Then

$$\Big|\sum_{n+1}^{\lambda n} s_\nu \nu^{-1} \sin \nu h\Big| < h \sum_{n+1}^{\lambda n} |s_\nu| < hk\epsilon^2(\lambda - 1)n = \epsilon,$$

and $\lambda n h = nh + 1/\epsilon > 1/\epsilon$, hence $B_n(h) = O(\epsilon)$, for $n > n_0(\epsilon)$. This proves Theorem 2. The analogous theorem holds for summability $(R. 1)$; see [4], Theorem 5.

Note that the convergence of $\sum a_n$ is not a necessary condition for the uniform convergence of (1.1) in the interval $(0, \pi)$. This is seen from the example in which s_n is 1 or 0 according as n is or is not a power of 2.

4. Summability (R_1) of Fourier series.

The following theorem is due to Hardy and Rogosinski [1, Theorem 3]:

In order that the Fourier series

$$(4.1)\qquad f(t) \sim \sum_{1}^{\infty} a_n \cos nt$$

should be summable (R_1) to 0 for $t = 0$, it is necessary and sufficient that the Cauchy integral

$$I(h) = \int_{\to 0}^{\pi} t^{-1} dt \int_{|h-t|}^{h+t} f(u) \cot(u/2)\, du$$

should converge for small h and tend to 0 with h. The upper limit may be replaced by any positive $\delta < \pi$. Hardy and Rogosinski have constructed a class of Fourier series which are nowhere summable (R_1). It is all the more remarkable that a Fourier series is summable (R_1) at each point of continuity.

For the proof we may assume that $t = 0$ is a point of continuity and that $f(0) = 0$. We write

$$I(h) = \int_{\to 0}^{h} + \int_{h}^{h^{\frac{1}{2}}} + \int_{h^{\frac{1}{2}}}^{\pi} = I_1 + I_2 + I_3,$$

say. Now

$$I_1(h) = o\left(\int_{\to 0}^{h} t^{-1} dt \int_{h-t}^{h+t} u^{-1} du\right) = o\left(\int_{\to 0}^{h} t^{-1} \log(h+t)/(h-t)\, dt\right);$$

the transformation $(h+t)/(h-t) = 1 + x$, $t = hx/(2+x)$, yields

$$(4.2) \qquad I_1(h) = o\left(\int_0^{\infty} x(2+x)^{-1} \log(1+x)\, dx\right) = o(1).$$

Next

$$I_2(h) = o \int_h^{h^{\frac{1}{2}}} t^{-1} dt \int_{t-h}^{t+h} u^{-1} du = o \int_h^{h^{\frac{1}{2}}} t^{-1} \log(t+h)/(t-h)\, dt,$$

and the transformation $(t+h)/(t-h) = 1 + x$, $t = h(2+x)/x$, yields

$$(4.3) \qquad I_2(h) = o \int_{2h/h^{\frac{1}{2}}-h}^{\infty} [x(2+x)]^{-1} \log(1+x)\, dx = o(1).$$

Finally

$$I_3(h) = O \int_{h^{\frac{1}{2}}}^{\pi} t^{-1} dt \int_{t-h}^{t+h} u^{-1} f(u)\, du$$

$$= O \int_{h^{\frac{1}{2}}}^{\pi} [t(t-h)]^{-1} [f_1(t+h) - f_1(t-h)]\, dt,$$

if $f_1(t) = \int_0^t | f(u) |\, du$.

Now $\int_{h^{\frac{1}{2}}}^{\pi} f_1(t-h)/[t(t-h)]\, dt = \int_{h^{\frac{1}{2}}-2h}^{\pi-2h} f_1(t+h)/[(t+2h)(t+h)]\, dt,$

so that

$$I_3(h) = O\left(\int_{h^{\frac12}-2h}^{\pi-2h} f_1(t+h)\{[t(t-h)]^{-1} - [(t+2h)(t+h)]^{-1}\}dt\right.$$

$\left. + \int_{\pi-2h}^{\pi} f_1(t+h)/[t(t-h)]\,dt\right)$. Furthermore $\int_{\pi-2h}^{\pi} f_1(t+h)/[t(t-h)]\,dt$ $= O\int_{\pi-2h}^{\pi}(t-h)^{-1}dt = O(h) = o(1)$, and

$$\int_{h^{\frac12}-2h}^{\pi-2h} f_1(t+h)\{[t(t-h)]^{-1} - [(t+2h)(t+h)]^{-1}\}dt$$

$$= 2h\int_{h^{\frac12}-h}^{\pi-h} f_1(t)(2t-h)/[t(t-h)(t+h)(t-2h)]\,dt$$

$$= O\left(h\int_{h^{\frac12}-h}^{\pi-h}(2t-h)/[(t-h)(t+h)(t-2h)]\,dt\right)$$

$$= O\left(\int_{h^{\frac12}}^{\pi}(2t-3h)h/[(t-2h)t(t-3h)]\,dt\right)$$

$$= O\left(\int_{h^{\frac12}}^{\pi} h/(t-3h)^2dt\right)$$

$= O(h/(h^{\frac12}-3h)) = o(1)$. Thus $I_3(h) = o(1)$, $h \to 0$. Hence, the estimates (4.2)-(4.3) prove the following theorem.

THEOREM 3. *The Fourier series of a function $f(t)$ is summable (R_1) at each point of continuity of $f(t)$.*

5. The Lebesgue constants. Employing well known results, Theorem 3 also follows if we prove that the set of Lebesgue constants corresponding to the transform (R_1) is bounded. We assume that $f(t)$ is integrable and that $|f(t)| \leq 1$, $0 \leq t \leq \pi$. Let

$$s_n = s_n(\theta) = \sum_1^n a_\nu \cos \nu\theta = (1/2)a_n \cos n\theta + s_n^*(\theta).$$

Then $s_n^*(\theta) = 1/\pi\int_0^\pi \phi(t)\cos(t/2)\sin nt\,dt$, where

$$\phi(t) = (1/2)[f(\theta+t) + f(\theta-t)].$$

Because of periodicity, we may restrict ourselves, as usual to $\theta = 0$. Now

$$\sum_1^\infty s_n n^{-1}\sin nh = (1/2)\sum_1^\infty a_n n^{-1}\sin nh + \sum_1^\infty s_n^* n^{-1}\sin nh;$$

here $\left|\sum_1^\infty a_n n^{-1}\sin nh\right| = \left|\int_0^h f(t)\,dt\right| \leq h/2 \leq \pi/2$ for $0 < h < \pi$. Next, from [1], § 3.1,

$$\sum_1^\infty s_n^* n^{-1}\sin nh = 1/\pi\int_0^\pi f(t)\cot(t/2)\sum_1^\infty n^{-1}\sin nt\sin nh\,dt,$$

where, if $0 < t < \pi$, $0 < h < \pi$,

$$\sum_1^\infty n^{-1} \sin nt \sin nh = (1/2) \log [\sin (1/2)(t+h)]/[\sin (1/2) \,|\, t-h \,|] > 0.$$

It now follows that $| \sum_1^\infty s_n^* n^{-1} \sin nh |$

$$\begin{aligned} &\leq (1/2\pi) \int_0^\pi \cot(t/2) \log[\sin(1/2)(t+h)]/[\sin(1/2)\,|\,t-h\,|]dt \\ &= (1/\pi) \int_0^\pi \cot(t/2) \sum_1^\infty n^{-1} \sin nt \sin nh \, dt \\ &= (1/\pi) \sum_1^\infty n^{-1} \sin nh \int_0^\pi \cot(t/2) \sin nt \, dt = \sum_1^\infty n^{-1} \sin nh \end{aligned}$$

$= (\pi - h)/2 \leq \pi$, termwise integration being legitimate, as in a similar case in [1], p. 178.

This proves that the Lebesgue constants are uniformly bounded.

6. Other conditions for summability (R_1).

THEOREM 4. *If*

$$(6.1) \qquad s_n - s = o(1/\log n),$$

and if for some $\delta > 0$

$$(6.2) \qquad \sum_n^{2n} (|\, a_\nu \,| - a) = O(n^{1-\delta}),$$

then $\sum a_n$ *is summable* (R_1).

Summability $(R, 1)$ under the same assumptions was proved in [3], p. 578.

We may assume that $s = 0$ and $\delta < 1$; we write

$$(6.3) \qquad \sum_1^\infty s_\nu \nu^{-1} \sin \nu h = \sum_1^n + \sum_{n+1}^\infty = C_n(h) + D_n(h),$$

say. It follows from Lemma 2 that $\sum \nu^{-1} \,|\, a_\nu \,| < \infty$; from (6.1) and (2.4) for fixed h, $s_n \rho_{n+1} \to 0$, and from (2.3) follows the convergence of (6.3).

We choose $n = (1/h)$, $0 < h < 1/2$, then, in view of (6.1),

$$(6.4) \qquad C_n(h) = O(h \sum_{\nu h \leq 1} |\, s_\nu \,|) = O[1/\log(1/h)] = o(1), \qquad h \to 0.$$

We next choose an r such that $r\delta > 1$; to any $h < 1/4$ choose $k_n = [h^{-r}] \leq h^{-r}$. Now

$$\begin{aligned} (6.5) \quad \sum_{n+1}^{k_n} s_\nu \nu^{-1} \sin \nu h &= o(\sum_{n+1}^{k_n} 1/\nu \log \nu) = o(\log\log k_n - \log\log n) \\ &= o(\log(\log k_n)/(\log n)) = o(\log r \log(1/h)/\log(1/h)) \end{aligned}$$

$= \log r \cdot o(1)$. Finally, employing Lemma 2' and formulas (2.4), (2.5), we have

$$(6.6) \qquad \sum_{k_n}^{\infty} s_\nu \nu^{-1} \sin \nu h = h^{-1} O(k_n^{-\delta}) = h^{-1} O(h^{r\delta}) = O(h^{r\delta - 1}) = o(1).$$

Combining (6.3)-(6.6) yields Theorem 4.

7. A $(C, r) \rightarrow (R_1)$ theorem.

THEOREM 5. *If* $\sum a_n$ *is summable* $(C, 1 - \alpha)$ *for some positive* $\alpha < 1$ *and if*

$$(7.1) \qquad \sigma_n \equiv \sum_1^n | s_\nu^{-\alpha} | = O(n^{1-\alpha}),$$

then the series is summable (R_1).

The corresponding theorem holds for summability $(R, 1)$; see [5], Theorem A. For closely related theorems see [2].

The Cesàro sums s_n^r are defined by

$$(7.2) \qquad s_n^r = \sum_1^n \gamma_{n-\nu}^r a_\nu, \text{ where } \gamma_n^r = \binom{n + r}{n}.$$

We have $\gamma_n^r - \gamma_{n-1}^r = \gamma_n^{r-1}$,

$$(7.3) \qquad s_n^r = \sum_1^n \gamma_{n-\nu}^{\epsilon - 1} s_\nu^{r - \epsilon}, \qquad s_n^0 = s_n,$$

$$(7.4) \qquad \gamma_n^r \sim n^r / \Gamma(r + 1), \qquad n \rightarrow \infty, r \neq -1, -2, \cdots.$$

By asumption, $s_n^{1-\alpha} / \gamma_n^{1-\alpha} \rightarrow s$, $n \rightarrow \infty$; we may assume $s = 0$, so that, in view of (7.4)

$$(7.5) \qquad n^{\alpha - 1} s_n^{1-\alpha} \rightarrow 0, \qquad n \rightarrow \infty.$$

From (7.3), for $r = 0$ and $\epsilon = \alpha - 1$, we have $s_n = \sum_1^n \gamma_{n-\nu}^{\alpha - 2} s_\nu^{1-\alpha}$; hence, from (7.4) and (7.5), $s_n = O(n^{1-\alpha}) = o(n)$. Thus, by Lemma 1, we need only to prove summability $(\bar{R}_1)$ to zero.

We have $a_n = \sum_1^n \gamma_{n-\nu}^{\alpha - 2} s_\nu^{-\alpha}$, hence

$$(7.6) \qquad \sum_1^\infty a_n \rho_n(h) = \sum_1^\infty \rho_n(h) \sum_{\nu=1}^n \gamma_{n-\nu}^{\alpha-2} s_\nu^{-\alpha} = \sum_{\nu=1}^\infty s_\nu^{-\alpha} \sum_{n=\nu}^\infty \gamma_{n-\nu}^{\alpha-2} \rho_n(h);$$

the interchange of summation is justified if the double series is absolutely convergent. Now, from (2.4) and (7.4),

$$\sum_{n \geqq \nu} | \gamma_{n-\nu}^{\alpha-2} \rho_n(h) | = O \sum_{n \geqq \nu} n^{-1} (n - \nu + 1)^{-2}$$

$= O(\nu^{-1})$ for $\alpha < 1$. Furthermore

$$\sum_1^m \nu^{-1} \mid s_\nu^{-\alpha} \mid = m^{-1}\sigma_m + \sum_1^{m-1} \sigma_\nu/\nu(\nu+1) = O(m^{-\alpha}) + O\sum_1^m \nu^{1-\alpha}/\nu(\nu+1)$$

$= O(1)$ as $0 < \alpha < 1$; this proves (7.6).

Summation by parts gives

$$\sum_{n=\nu}^m \gamma_{n-\nu}{}^{\alpha-2}\rho_n(h) = \gamma_{m-\nu}{}^{\alpha-1}\rho_m(h) + \sum_{n=\nu}^{m-1} \gamma_{n-\nu}{}^{\alpha-1}(\rho_n - \rho_{n+1})$$

$= O(m^{\alpha-2}) + \sum_{n=\nu}^{m-1} \gamma_{n-\nu}{}^{\alpha-1} n^{-1} \sin nh$, $\alpha < 2$, hence

$$\sum_{n=\nu}^\infty \gamma_{n-\nu}{}^{\alpha-2}\rho_n(h) = \sum_{n=\nu}^\infty \gamma_{n-\nu}{}^{\alpha-1} n^{-1} \sin nh \equiv \Delta^{-\alpha}\nu^{-1}\sin \nu h.$$

Now, from (7.6)

$$\sum_1^\infty a_n\rho_n(h) = \sum_{\nu=1}^\infty s_\nu{}^{-\alpha}\Delta^{-\alpha}\nu^{-1}\sin \nu h = \sum_{\nu h \leqq \lambda} + \sum_{\nu h > \lambda} \tag{7.7}$$

$= \phi_1(h) + \phi_2(h)$, say, where $\lambda > 1$ is fixed.

We wish to show that $\phi_1(h) \to 0$ and $\phi_2(h) \to 0$, as $h \to 0$. To this end we first estimate $\Delta^r\nu^{-1}\sin \nu h$ for $r > -1$. By definition,

$$\Delta^r\nu^{-1}\sin \nu h = \sum_{n=\nu}^\infty \gamma_{n-\nu}{}^{-r-1} n^{-1}\sin nh, \qquad r > -1.$$

Now $n^{-1}\sin nh = h/2 \int_{-1}^1 e^{inhx}dx$, hence

$$\Delta^r\nu^{-1}\sin \nu h = (h/2)\sum_\nu^\infty \gamma_{n-\nu}{}^{-r-1}\int_{-1}^1 e^{inhx}dx = h/2 \int_{-1}^1 \sum_\nu^\infty \gamma_{n-\nu}{}^{-r-1}e^{inhx}dx,$$

interchange of summation and integration being legitimate. But

$$\sum_{n=\nu}^\infty \gamma_{n-\nu}{}^{-r-1}e^{inhx} = \sum_{k=0}^\infty k^{-r-1}e^{i(\nu+k)hx}$$

$= e^{i\nu hx}(1 - e^{ihx})^r$, hence

$$\Delta^r\nu^{-1}\sin \nu h = h/2 \int_{-1}^1 e^{i\nu hx}(1 - e^{ihx})^r dx$$

It follows that

$$\mid \Delta^r\nu^{-1}\sin \nu h \mid < h \int_{-1}^1 \mid 1 - e^{ihx} \mid^r dx = O(h^{1+r}). \tag{7.8}$$

We now assume $0 < h < 1/2$, and write

$$[\lambda h^{-1}] = n; \tag{7.9}$$

thus

$$\phi_1(h) = \sum_{\nu=1}^{n} s_\nu^{-\alpha}\Delta^{-\alpha}\nu^{-1}\sin \nu h = s_n^{1-\alpha}\Delta^{-\alpha}n^{-1}\sin nh + \sum_{1}^{n-1} s_\nu^{1-\alpha}\Delta^{1-\alpha}\nu^{-1}\sin \nu h.$$

Employing (7.5) and (7.8) we get

$$\phi_1(h) = o(n^{1-\alpha}h^{1-\alpha}) + o(\sum_{1}^{n-1}\nu^{1-\alpha}h^{2-\alpha})$$

$= o(nh)^{1-\alpha} + o(nh)^{2-\alpha}$, and in view of (7.9)

$$\phi_1(h) = o(\lambda^{1-\alpha}) + o(\lambda^{2-\alpha}) = o(\lambda^{2-\alpha}). \tag{7.10}$$

To estimate $\phi_2(h)$, consider again $\Delta^{-\alpha}\nu^{-1}\sin \nu h$, and write

$$\Delta^{-\alpha}\nu^{-1}\sin \nu h = \sum_{k=\nu}^{\nu+\eta}\gamma_{k-\nu}{}^{\alpha-1}k^{-1}\sin kh + \sum_{k=\nu+\eta+1}^{\infty}\gamma_{k-\nu}{}^{\alpha-1}k^{-1}\sin kh$$

$= \delta_1(h) + \delta_2(h)$, say, where $\eta = [h^{-1}]$. Now

$$\delta_1(h) = O\sum_{k=\nu}^{\nu+\eta}(k-\nu+1)^{\alpha-1}k^{-1} = O(\nu^{-1}\eta^{\alpha}) = O(\nu^{-1}h^{-\alpha}),$$

and, summing by parts,

$$\delta_2(h) = -\gamma_{\eta+1}{}^{\alpha-1}\rho_{\nu+\eta}(h) + \sum_{k=\nu+\eta+1}^{\infty}\gamma_{k-\nu}{}^{\alpha-2}\rho_k(h)$$

$$= O(\eta^{\alpha-1}[h(\nu+\eta)]^{-1}) + O(\sum_{k=\nu+\eta+1}^{\infty}(k-\nu+1)^{\alpha-2}(kh)^{-1})$$

$$= O(\eta^{\alpha-1}[h(\nu+\eta)]^{-1}) + O(\eta^{\alpha-1}[h(\nu+\eta)]^{-1})$$

$= O(\eta^{\alpha-1}(\nu h)^{-1}) = O(\nu^{-1}h^{-\alpha})$. Thus $\Delta^{-\alpha}\nu^{-1}\sin \nu h = O(\nu^{-1}h^{-\alpha})$. Now

$$\phi_2(h) = O(\sum_{\nu h > \lambda} | s_\nu^{-\alpha} | \nu^{-1}h^{-\alpha}),$$

furthermore, employing (7.1),

$$\sum_{k}^{\infty} | s_\nu^{-\alpha} | \nu^{-1} = k^{-1}\sigma_{k-1} + \sum_{k}^{\infty}\sigma_\nu/[\nu(\nu+1)]$$

$= O(\sum_{k}^{\infty}\nu^{-1-\alpha}) = O(k^{-\alpha})$, hence

$$\phi_2(h) = O(h^{-\alpha}(\lambda/h)^{-\alpha}) = O(\lambda^{-\alpha}). \tag{7.11}$$

Combining (7.7), (7.10) and (7.11), we get $\sum_{n=1}^{\infty} a_n\rho_n(h) = o(\lambda^{2-\alpha} + O(\lambda^{-\alpha})$; thus, letting $h \to 0$,

$$\limsup_{h\to 0} | \sum_{1}^{\infty} a_n\rho_n(h) | \leq O(\lambda^{-\alpha}).$$

Here α is positive, $O(1)$ is independent of λ, and λ is arbitrarily large, hence $\sum a_n\rho_n(h) \to 0$ as $h \to 0$. This proves Theorem 5.

8. Connection with summability $(k, 1)$. The well known formula

$$\sum_1^n \nu^{-1}\sin \nu t = -(t/2) + \int_0^t \{\sin[n+(1/2)]x/[2\sin(x/2)]\}dx$$

yields

$$\sum_1^{n-1} \nu^{-1}\sin \nu t + (1/2)n^{-1}\sin nt = -(t/2)+(1/2)\int_0^t \sin nx\cot(x/2)\,dx,$$

hence, for $0 < t < \pi$,

$$\rho_n(t) = \pi/2 + (1/2)n^{-1}\sin nt - (1/2)\int_0^t \sin nx\cot(x/2)\,dx.$$

Furthermore $\int_0^t x^{-1}\sin nx dx = \int_0^{nt} y^{-1}\sin y dy = \pi/2 - \int_{nt}^{\infty} y^{-1}\sin y dy$, so that

$$\begin{aligned}\rho_n(t) &= \int_{nt}^{\infty} y^{-1}\sin y dy + (1/2)n^{-1}\sin nt + \int_0^t [x^{-1}-(1/2)\cot(x/2)]\sin nx dx\\ &= Si(nt)+(1/2)n^{-1}\sin nt + \int_0^{t/2}(y\sin y)^{-1}(\sin y - y\cos y)\sin 2ny dy.\end{aligned}$$

The formula $(d/dy)(y^{-1}-\cot y) = (\sin^2 y)^{-1} - y^{-2}$ and an integration by parts yield

$$\begin{aligned}\rho_n(t) &= Si(nt)+(1/2)n^{-1}\sin nt + (t^{-1}-(1/2)\cot(t/2))n^{-1}(1-\cos nt)\\ &\quad -(1/2)\int_0^{t/2}[(\sin^2 y)^{-1}-y^{-2}]n^{-1}(1-\cos 2ny)\,dy.\end{aligned}$$

We now have

$$\begin{aligned}(8.1)\qquad \sum_1^n a_\nu\rho_\nu(t) &= \sum_1^n a_\nu Si(\nu t) + (1/2)\sum_1^n \nu^{-1}a_\nu \sin \nu t\\ &\quad + ((t^{-1}-(1/2)\cot(t/2))\sum \nu^{-1}a_\nu(1-\cos\nu t)\\ &\quad -(1/2)\int_0^{t/2}((\sin^2 y)^{-1}-y^{-2})\sum_1^n \nu^{-1}a_\nu(1-\cos 2\nu y)dy.\end{aligned}$$

We now suppose that both (4.1) and its conjugate

$$(8.2)\qquad \sum_1^\infty a_n \sin nt \sim \tilde{f}(t)$$

are Fourier series. It then follows from [1], Theorem 7, that $\sum a_n$ is summable (R_1) if and only if the Cauchy integral

$$(8.3)\qquad \int_{\to 0}^{\pi} \tilde{f}(t)/t\,dt$$

exists. We give here a simpler proof, based on formula (8.1). We now have

$$\sum_1^\infty n^{-1}a_n \sin nt = \int_0^t f(x)\,dx, \qquad \sum_1^\infty n^{-1}a_n(1-\cos nt) = \int_0^t \tilde{f}(x)\,dx,$$

and, from (8. 1),

$$\sum_1^\infty a_\nu\{\rho_\nu(t) - Si(\nu t)\} = 1/2 \int_0^t f(x)\,dx + (t^{-1} - (1/2)\cot(t/2)) \int_0^t \tilde{f}(x)\,dx$$
$$- 1/2 \int_0^{t/2} \{(\sin^2 y)^{-1} - y^{-2}\} \int_0^{2y} f(x)\,dx\,dy.$$

Here each term on the right tends to 0, as $t \to 0$. Thus the series $\sum_1^\infty a_\nu\rho_\nu(t)$ and $\sum a_\nu Si(\nu t)$ are either both convergent or both divergent. Now from (8. 2)

$$\sum_1^\infty a_\nu Si(\nu t) = \int_t^\infty \tilde{f}(x)/x\,dx, \qquad t > 0,$$

and the series converges for all $t > 0$; hence $\sum a_\nu\rho_\nu(t)$ is convergent for all $t > 0$. Furthermore $s_n = o(n)$, hence by Lemma 1, the series (1. 1) is convergent. It is now clear that summability (R_1) implies (8. 3) and conversely. We call the series $\sum a_n$ summable $(k, 1)$ to s if the series $\sum a_n Si(nt)$ converges in some interval $0 < t < t_0$, and if $2/\pi \sum a_n Si(nt) \to s$, as $t \to 0$. We have just shown that under the assumptions (4. 1) and (8. 2) the summability methods (R_1) and $(k, 1)$ are equivalent.

9. Gibbs phenomenon. In addition to the regularity of the corresponding Lebesgue constants, the transform (R_1) presents no Gibbs phenomenon. Consider the transform of the series $\sum_1^\infty n^{-1}\sin nt$, so that

$$s_n = -(t/2) + (1/2)n^{-1}\sin nt + (1/2)\int_0^t \cot(x/2)\sin nx\,dx;$$

thus the transform is

$$2/\pi \sum_1^\infty n^{-1}s_n \sin nh = -(t/\pi)\sum_1^\infty n^{-1}\sin nh + 1/\pi \sum_1^\infty n^{-2}\sin nt \sin nh$$
$$+ 1/\pi \int_0^t \cot(x/2)\left(\sum_1^\infty n^{-1}\sin nx \sin nh\right)dx.$$

It is easy to see that the transform sums the series to its proper value $(\pi - t)/2$ for $0 < t < \pi$. We wish to determine

$$\limsup_{t\to 0, h\to 0} 4/\pi^2 \sum n^{-1}s_n(t)\sin nh = \lambda,$$

which is called the Gibbs ratio of the transform (R_1); it is clear that $\lambda \geq 1$.

If $\lambda = 1$ then we say that the transform presents no Gibbs phenomenon. The first term on the right of (9.1) evidently tends to zero. The second term is in absolute value less than

$$1/2\pi \sum_1^\infty n^{-2}(\sin^2 nt + \sin^2 nh) = 1/2\pi\{t(\pi - t)/2 + h(\pi - h)/2\},$$

hence tends to zero. Finally the last term is positive and less than (see section 5)

$$1/\pi \int_0^\pi \mathrm{cot})x/2)\, (\sum_1^\infty n^{-1} \sin nx \sin nh)\, dx = (\pi - h)/2 < \pi/2,$$

which proves our assertion.

NATIONAL BUREAU OF STANDARDS AT LOS ANGELES
AND
UNIVERSITY OF CINCINNATI

BIBLIOGRAPHY.

[1] G. H. Hardy and W. W. Rogosinski, "Notes on Fourier Series (V): Summability (R_1)," *Proceedings of the Cambridge Philosophical Society*, vol. 45 (1949), pp. 173-185.

[2] N. Obreschkoff, "Über das Riemannsche Summierungsverfahren," *Mathematische Zeitschrift*, vol. 48 (1942-43), pp. 441-454.

[3] Otto Szász, "On convergence and summability of trigonometric series," *American Journal of Mathematics*, vol. 64 (1942), pp. 575-591.

[4] ———, "On Abel and Lebesgue summability," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 885-893.

[5] ———, "On Lebesgue summability and its generalization to integrals," *American Journal of Mathematics*, vol. 67 (1945), pp. 389-396.

LINEAR DIFFERENTIAL EQUATIONS WITH ALMOST PERIODIC OR LAPLACE TRANSFORM COEFFICIENTS.*

By Calvin R. Putnam and Aurel Wintner.

1. This paper consists of two parts, dealing with an n-th order linear differential equation possessing coefficients all of which either are uniformly almost periodic functions or are Laplace transforms. The proof, given in part II, of the theorem dealing with Laplace transforms will be somewhat abbreviated owing to its similarity with the proof, which will be given in detail in part I, of the theorem in the almost periodic case.

PART I.

2. Let

$$f_0(t), f_1(t), \cdots, f_{n-1}(t), \qquad (-\infty < t < \infty), \tag{1}$$

be n uniformly almost periodic functions of the real variable t, having the property that the frequencies $\lambda_1, \lambda_2, \cdots$ occurring in the Fourier expansion,

$$\sum_k a_{kj} e^{-i\lambda_k t}, \qquad j = 0, 1, \cdots, n-1, \tag{2}$$

of any of the functions (1) have a positive lower bound; so that, without loss of generality,

$$\lambda_k \geqq 1, \tag{3}$$

if the unit of length on the t-axis is suitably chosen. (Since the functions $f_k(t)$ are considered on the entire t-axis, the choice of the minus sign in (2) is, at present, obviously immaterial; it is inserted, however, for the sake of later convenience.)

Theorem I. The homogeneous, linear differential equation

$$\sum_{h=0}^{n} f_h(t)\, d^h x/dt^h = 0, \tag{4}$$

where $-\infty < t < \infty$ and the coefficients (1) are uniformly almost periodic on $-\infty < t < \infty$ with Fourier expansions (2) satisfying (3) and

$$f_n(t) \equiv 1, \tag{5}$$

---

* Received December 15, 1950.

must have a non-trivial ($\not\equiv 0$) solution $x = x_1(t)$ which is bounded, as $t \to \pm\infty$, and is, to a constant factor, the only bounded solution of (4);

this solution is uniformly almost periodic;

if c denotes the constant term of the Fourier expansion of $x_1(t)$, then $c \neq 0$ (so that, since $x_1(t)$ can be multiplied by any constant, c can be assigned as an integration constant determining $x_1(t)$, the trivial solution $x(t) \equiv 0$ being determined by $c = 0$);

if the constant term of the Fourier expansion of $x_1(t)$ is subtracted from $x_1(t)$, then the resulting uniformly almost periodic function, $x_1(t) - c$, has frequencies all of which are linear combinations, with positive integral coefficients, of the frequencies of the coefficient functions (1) (so that, in particular, all non-vanishing frequencies of $x_1(t)$ must satisfy the inequality (3), the latter being assumed of all frequencies of the functions (1));

the frequency 0 of $x_1(t)$, represented by the occurrence of the term $c \neq 0$ in the Fourier expansion of $x_1(t)$, can be thought of as being due to the presence of the coefficient function (5) (which is uniformly almost periodic with 0 as its only frequency);

finally, the general solution of (4) can be described as follows:

(4) has a system of n linearly independent solutions $x = x^1(t), \cdots, x^n(t)$, which can be chosen so as to possess the recursive structure

$$x^1(t) = x_1(t), x^2(t) = tx^1(t) + x_2(t), \cdots, x^n(t) = tx^{n-1}(t) + x_n(t), \tag{6}$$

that is,

$$x^k(t) = \sum_{h=0}^{k-1} t^h x_{k-h}(t), \qquad k = 1, 2, \cdots, n, \tag{6 bis}$$

where $x_1(t)$ is the uniformly almost periodic function singled out above, and the remaining $n - 1$ functions $x_2(t), \cdots, x_n(t)$ have the structure

$$x_k(t) = \sum_{h=0}^{k-2} t^h x_{kh}(t), \qquad k = 2, 3, \cdots, n, \tag{7}$$

where the functions x_{kh} are uniformly almost periodic with Fourier expansions satisfying (2) and (3) (and with frequencies all of which are linear combinations, with positive integral coefficients, of the frequencies of the coefficient functions (1)).

Corresponding results hold when (4) is replaced by the inhomogeneous differential equation

$$\sum_{h=0}^{n} f_h(t) d^h x/dt^h = g(t), \tag{8}$$

if $g(t)$ is uniformly almost periodic and satisfies (3), as do the coefficient functions (1). It will be clear from the results of Section 5 below that (8) has a particular solution which is almost periodic and has a Fourier expansion (2) satisfying (3) with frequencies which are linear combinations, with positive integral coefficients, of the frequencies of the coefficient functions (1) and of $g(t)$. The general solution of (8) is obtained, of course, by adding to this particular solution the general solution of the linear homogeneous differential equation (4), as furnished by (6).

The truth of the assertion of the part of the theorem as to the frequencies of certain functions being linear combinations, with positive integral coefficients, of the frequencies of the coefficient functions, will be clear from the proof.

All of this parallels the results of [4], where $n=2$, in the same way as [2] parallels [3].

In order to prove Theorem I, it will be convenient to transcribe the statement and terminology of this theorem in such a manner that the real line $-\infty < t < \infty$ is replaced by the complex closed half-plane $\sigma \geqq 0$, where $s = \sigma + it$. For this purpose (cf. [2], p. 860), a function $f(s)$ will be called uniformly almost periodic, in the half-plane $\sigma \geqq 0$, if $f(s)$ is regular and, in the usual sense (cf. [1], p. 141), is uniformly almost periodic in the open half-plane $\sigma > 0$, is continuous on the closed half-plane $\sigma \geqq 0$, and $f(it)$ $(=f(s)$ on the boundary line $\sigma = 0)$ is uniformly almost periodic for $-\infty < t < \infty$. If the variable it is replaced by s in each of the series (2), it follows from the theorem of [1], pp. 150-151, that each of the new series

$$(2') \qquad \sum_k a_{kj} e^{-\lambda_k s}$$

is the Fourier expansion of a corresponding function

$$(1') \qquad F_0(s), F_1(s), \cdots, F_{n-1}(s),$$

uniformly almost periodic for $\sigma \geqq 0$, and such that $F_h(it) = f_h(t)$, $h = 0, 1, \cdots, n-1$. Define $F_n(s)$ by

$$(5') \qquad F_n(s) \equiv 1.$$

Let (6′), (6′ bis), (7′) denote the relations which result by replacing t by s in (6), (6 bis), (7), respectively. Correspondingly, replace (4) by

$$(4') \qquad \sum_{h=0}^{n} F_h(s)\, d^h x/ds^h = 0$$

and (8) by

$$(8') \qquad \sum_{h=0}^{n} F_h(s)\, d^h x/ds^h = G(s),$$

where now $x = x(s)$, the functions $F_h(s)$ are those defined above, and $G(s)$ is obtained by replacing it by s in the function $g(t) = G(it)$. If the statement of Theorem I is altered so as to correspond to the primed relations just introduced, there results, for the complex half-plane $\sigma \geqq 0$, a theorem, to be called Theorem I′, which is similar to Theorem I for the real line $-\infty < t < \infty$. That the assumptions of Theorem I′, where uniform almost periodicity refers to the half-plane $\sigma \geqq 0$, are implied by those of Theorem I, is a consequence of the preceding paragraph. Let $s = it$ (that is, $\sigma = 0$), so that the equation (4′) [or (8′)], if multiplied by i^n, becomes

$$\sum_{h=0}^{n} i^{n-h} f_h(t)\, d^h x/dt^h = 0 \quad [\text{or} = i^n g(t)].$$

It is clear that the functions $i^{h-n} f_h(t)$ and $g(t)/i^n$ have the same structure as the respective functions $f_h(t)$ and $g(t)$. Moreover, if any solution x^k of (4) is multiplied by a non-vanishing constant, the essential properties of x^k as regard to its structure and to its being a solution of (4) are unchanged.

It is understood that all derivatives dy/ds, considered above or in the sequel, are to be interpreted on the boundary line $\sigma = 0$ in the "one-sided" sense. Thus, $(dy/ds)_{s=s_0} = \lim\, (f(s) - f(s_0))/(s - s_0)$, where $s_0 = it_0$ is fixed and $s = \sigma + it$, $\sigma \geqq 0$. In addition, it is to be noted that, on the half-plane $\sigma \geqq 0$, the linear independence of the solutions $x^k(s)$, considered in Theorem I′ (corresponding to (6) for Theorem I), implies their linear independence on $\sigma = 0$.

The above remarks, together with an inspection of the statement of Theorem I′ when $\sigma = 0$, make it clear that Theorem I is a corollary of Theorem I′. The proof of Theorem I′ will be given in Sections 4, 5 and 6.

3. Two lemmas will first be proved.

LEMMA 1. *On the half-plane* $\sigma \geqq 0$, *let* $f(s)$ *be a uniformly almost periodic function with a Fourier expansion*

$$f(s) \sim \sum_k a_k e^{-\mu_k s}, \qquad \text{where } \mu_m \geqq \nu > 0, \tag{9}$$

and let n denote an arbitrary positive integer. Then there exist a constant c such that $|f(s)| \leqq ce^{-\nu\sigma}$ *for* $\sigma \geqq 0$, *and* $n+1$ *functions* $f(s) = g_0(s), g_1(s), \cdots, g_n(s)$, *each uniformly almost periodic for* $\sigma \geqq 0$ *and possessing a Fourier expansion of the type* (9) (*with the same frequencies* μ_k *as for* $f(s)$), *such that*

$$dg_k/ds = g_{k-1} \ (\text{that is, } d^k g_k/ds^k = f) \text{ for } \sigma \geqq 0 \text{ and } k = 1, 2, \cdots, n.$$

Furthermore,

$$(10)\qquad |g_k| \leqq c\nu^{-k}e^{-\nu\sigma} \text{ for } \sigma \geqq 0, \qquad k = 0, 1, \cdots, n,$$

and hence, in particular,

$$(11)\qquad |g_k| \leqq c\nu^{-1}e^{-\nu\sigma} \text{ for } \sigma \geqq 0, \qquad k = 1, 2, \cdots, n.$$

It is easy to see that the function $f^*(s)$ defined in [2], p. 862, formula line (16), satisfies the equation $df^*/ds = f(s)$ (loc. cit., p. 861, formula line (14)) for $\sigma \geqq 0$, rather than merely for $\sigma > 0$. This fact, and the proof and successive applications of the Lemma, loc. cit., p. 861, readily imply Lemma 1 above.

LEMMA 2. *On the half-plane $\sigma \geqq 0$, let $h_0(s), h_1(s), \cdots, h_M(s)$ denote $M + 1$ uniformly almost periodic functions, each having a Fourier expansion of the type* (9). *Then, by Lemma* 1 *(cf.* (10) *for $k = 0$), there exists a constant c such that $|h_k(s)| \leqq ce^{-\nu\sigma}$ for $\sigma \geqq 0$ and $k = 0, 1, \cdots, M$. Furthermore, the differential equation*

$$(12)\qquad d^n x/ds^n = \sum_{k=0}^{M} s^k h_k(s), \qquad \sigma \geqq 0,$$

possesses a solution x such that

$$(13)\qquad d^l x/ds^l = \sum_{k=0}^{M} s^k x_{kl}, \qquad l = 0, 1, \cdots, n-1; \sigma \geqq 0,$$

where the functions x_{kl} are uniformly almost periodic for $\sigma \geqq 0$, with Fourier expansions satisfying (9) *(with frequencies which are contained in the set of frequencies of the functions h_k), and are such that*

$$(14)\qquad |x_{kl}| \leqq C_M\nu^{-1}ce^{-\nu\sigma}, \text{ for all } k \text{ and } l = 0, 1, \cdots, n-1,$$

where C_M denotes a constant depending only on M.

The proof of Lemma 2 depends on a number of partial integrations and on Lemma 1. If $[y]^{(m)}$ denotes a certain m-th primitive of a function y, a partial integration shows that

$$(15)\qquad [s^k h_k]^{(1)} = s^k g_1 - k[s^{k-1}g_1]^{(1)},$$

where the present h_k and g_1 are identified with the functions f and g_1 respectively, occurring in Lemma 1. Successive partial integrations applied to (15) and the use of Lemma 1 show that

$$[s^k h_k]^{(1)} = \sum_{h=0}^{k} s^h z_{kh},$$

where the functions z_{kh} are uniformly almost periodic for $\sigma \geqq 0$ and of type (9), and satisfy

$$| z_{kh} | \leqq k\,!\,\nu^{-1} c e^{-\nu\sigma}, \qquad \sigma \geqq 0.$$

Since $k \leqq M$, it is readily seen that

$$[\sum_{k=0}^{M} s^k h_k]^{(1)} = \sum_{h=0}^{M} s^h z_{kh1}, \text{ where } | z_{kh1} | \leqq (M+1)\,!\,\nu^{-1} c e^{-\nu\sigma},$$

holds for $k = 0, 1, \cdots, M$. (The properties of the functions z_{kh1} are described below.) It follows that

$$(16) \qquad [\sum_{n=0}^{M} s^k h_k]^{(m)} = \sum_{h=0}^{M} s^h z_{khm}, \qquad m = 1, 2, \cdots, n,$$

where the functions z_{khm} are uniformly almost periodic for $\sigma \geqq 0$, are of type (9), and satisfy

$$| z_{khm} | \leqq [(M+1)\,!]^m \nu^{-m} c e^{-\nu\sigma}, \qquad m = 1, 2, \cdots, n.$$

Relations (16) and (12) now imply (13) and (14) if x is defined by $x = [\sum_{h=0}^{M} s^k h_k]^{(n)}$ and $c_M = [(M+1)\,!]^n$. This completes the proof of Lemma 2.

4. *Proof of Theorem* I′. The existence of the function $x = x^1(s)$, corresponding to the $x^1(t)$ of Theorem I, satisfying the differential equation

$$(4') \qquad \sum_{h=0}^{n} F_h(s)\, d^h x / ds^h \doteq 0, \qquad \sigma \geqq 0,$$

will first be shown. The differential equations, determining the successive approximations $y_1, y_2, \cdots$ to the desired solution x^1, will be set up for $\sigma \geqq 0$ as follows:

$$(17) \qquad d^n y_{m+1} / ds^n = -\sum_{h=0}^{n-1} F_h(s)\, d^h y_m / ds^h, \qquad m = 0, 1, 2, \cdots,$$

where

$$(18) \qquad y_0(s) \equiv 1.$$

For $m = 0$, (17) becomes, in virtue of (18),

$$(19) \qquad d^n y_1 / ds^n = -F_0(s).$$

Let c denote a constant such that

$$(20) \qquad | F_k(s) | \leqq c e^{-\sigma}, \qquad \sigma \geqq 0,\ k = 0, 1, \cdots, n-1;$$

cf. Lemma 1. An application of Lemma 1 shows that (19) has a solution y_1

which, along with its first $n-1$ derivatives, is uniformly almost periodic for $\sigma \geqq 0$, with a Fourier expansion (2′) satisfying (3), and such that

$$(21) \qquad | d^l y_1/ds^l | \leqq ce^{-\sigma}, \qquad \sigma \geqq 1, l = 0, 1, \cdots, n-1.$$

(The above functions y_1 and $-F_0(s)$ correspond to g_n and $f.(s)$ in Lemma 1.) Suppose that a function y_m has been determined so that it and its first $n-1$ derivatives are uniformly almost periodic for $\sigma \geqq 0$ and possess Fourier expansions (2′), with exponents $\lambda_k \geqq m$, and are such that

$$(22) \qquad | d^l y_m/ds^l | \leqq (nce^{-\sigma})^m/m!, \qquad \sigma \geqq 0, l = 0, 1, \cdots, n-1.$$

(These conditions are clearly satisfied for $m = 1$.) Then, by (20) and (22),

$$\sum_{h=0}^{n-1} F_n(s) d^h y_m/ds^h | \leqq (nce^{-\sigma})^{m+1}/m!.$$

Another application of Lemma 1 shows that (17) possesses a solution y_{m+1} such that it and its first $n-1$ derivatives are uniformly almost periodic for $\sigma \geqq 0$ and satisfy

$$| d^l y_{m+1}/ds^l | \leqq (nce^{-\sigma})^{m+1}/(m+1)!, \qquad \sigma \geqq 0, l = 0, 1, \cdots, n-1.$$

Thus, (22) is satisfied if m is replaced by $m+1$. Consequently, by (18), (21) and (22) (the last being valid for $m = 0, 1, 2, \cdots$),

$$(24) \qquad \sum_{m=1}^{\infty} | d^l y_m/ds^l | \leqq CA(\sigma), \qquad \sigma \geqq 0, l = 0, 1, \cdots, n-1,$$

where C denotes a constant independent of l and s, and $A(\sigma) \rightarrow 0$ holds as $\sigma \rightarrow \infty$. It is seen from (24) and (17) that the inequality (24) holds also for $l = n$.

The preceding results, together with (17) and (18), show that the function x^1 defined by

$$(25) \qquad x^1(s) = 1 + \sum_{m=1}^{\infty} y_m(s)$$

is a solution of (4′). Since the series of (25) is uniformly convergent for $\sigma \geqq 0$ and since each term is uniformly almost periodic, it follows that $x^1(s)$ (together with its first n derivatives, cf. (24), (4′) and (5′)) is uniformly almost periodic for $\sigma \geqq 0$.* That x^1 is bounded and $\neq 0$, with 1 as its non-vanishing constant term, follows from (25) and (24), for $l=0$. The remaining properties of $x^1(s)$ concerning its non-vanishing exponents are clear from the above construction.

---

* In [2], p. 864, the matrix function $W(s)$ should be defined (in the terminology occurring there) by $W(s) = \sum_{0}^{\infty} V_m(s)$. Since $\sum_{0}^{\infty} | V'_m(s) |$ is uniformly convergent, W is the solution of the differential equation under consideration for which $W - E$ has no constant term.

5. Consider now the inhomogeneous equation

$$\sum_{h=0}^{n} F_h(s)\, d^h x/ds^h = g(s), \qquad \sigma \geqq 0, \tag{8''}$$

where $g(s)$ is not in general uniformly almost periodic but is defined by

$$g(s) = \sum_{k=0}^{M} s^k h_k, \tag{26}$$

where the functions $h_0(s), h_1(s), \cdots, h_M(s)$ of (26) are supposed to be uniformly almost periodic of type (2′) and (3). It will be shown that (8″) has a solution $x(s)$ such that

$$d^l x/ds^l = \sum_{k=0}^{M} s^k z_{kl}, \qquad l = 0, 1, \cdots, n-1, \tag{27}$$

where the functions z_{kl} are uniformly almost periodic for $\sigma \geqq 0$ and have Fourier expansions of type (2′) and (3).

Consider first the equation

$$d^n y_1/ds^n = g(s), \qquad \sigma \geqq 0. \tag{28}$$

According to Lemma 2, (28) possesses a solution y_1 such that

$$d^l y_1/ds^l = \sum_{h=0}^{M} s^h y_{1hl}, \qquad l = 0, 1, \cdots, n-1, \tag{29}$$

where the functions y_{1hl} are uniformly almost periodic for $\sigma \geqq 0$ and are of the type (2′) and (3). Consider the equation

$$d^n y_{m+1}/ds^n = -\sum_{l=0}^{n-1} F_l(s)\, d^l y_m/ds^l, \qquad m = 1, 2, \cdots, \tag{30}$$

where the functions $F_k(s)$ are given by (1′). If $m=1$, (30) becomes

$$d^n y_2/ds^n = -\sum_{l=0}^{n-1} F_l(s)\, d^l y_1/ds^l, \tag{31}$$

where y_1 is defined by (29) for $l=0$. By (29) and the properties of the functions (1′), (31) can be written as

$$d^n y_2/ds^n = \sum_{h=0}^{M} s^h g_h, \tag{32}$$

where the functions g_h are uniformly almost periodic for $\sigma \geqq 0$ and are of type (9) for $\nu = 2$. Hence, by Lemma 1, there exists a constant c satisfying

$$(33) \qquad | g_k | \leqq ce^{-2\sigma} \text{ for } \sigma \geqq 0.$$

It follows from Lemma 2 that (32) possesses a solution y_2 such that

$$d^l y_2/ds^l = \sum_{h=0}^{M} s^h y_{2hl}, \qquad l = 0, 1, \cdots, n-1,$$

where the functions y_{2hl} are uniformly almost periodic for $\sigma \geqq 0$ and satisfy

$$| y_{2hl} | \leqq C_M ce^{-2\sigma}/2, \qquad \sigma \geqq 0, l = 0, 1, \cdots, n-1,$$

(cf. (14)). Suppose that y_m, where $m \geqq 2$, has been determined such that

$$(34) \qquad d^l y_m/ds^l = \sum_{h=0}^{M} s^h y_{mhl}, \qquad l = 0, 1, \cdots, n-1,$$

where the y_{mhl} are uniformly almost periodic for $\sigma \geqq 0$ with Fourier expansions of type (9) for $\nu = m$ and are such that

$$(35) \qquad | y_{mhl} | \leqq (C_M nc)^{m-1} e^{-m\sigma}/m!, \qquad \sigma \geqq 1, l = 0, 1, \cdots, n-1.$$

Then

$$(36) \qquad -\sum_{l=0}^{n-1} F_l(s) d^l y_m/ds^l = \sum_{h=0}^{M} s^h [-\sum_{l=0}^{n-1} F_l(s) y_{mhl}],$$

where, by the properties of the functions (1′) and the functions y_{mhl}, the functions $[\cdot \cdot \cdot]$ of (36), where $h = 0, \cdots, M$, are uniformly almost periodic for $\sigma \geqq 0$, are of type (9) for $\nu = m + 1$, and satisfy

$$| \sum_{l=0}^{n-1} F_l(s) y_{mhl} | \leqq (nc)^m C_M{}^{m-1} e^{-(m+1)\sigma}/m!, \qquad \sigma \geqq 0.$$

An application of Lemma 2 shows that (30) possesses a solution y_{m+1} such that (34) and (35) remain valid if m is replaced by $m + 1$. The preceding results, together with (28) and (30), show that the function $x(s)$ defined by

$$(37) \qquad x(s) = \sum_{m=1}^{\infty} y_m(s)$$

satisfies (8″), where $g(s)$ is defined by (26). It is clear from the properties of the function $y_1, y_2, \cdots$ that $x(s)$ and its first $n-1$ derivatives have representations of the form (27), where the z_{kl} are uniformly almost periodic of type (2′) and (3).

6. Consider finally the equation (4′). Its solution $x = x^1(s)$ (corresponding to the function $x^1(t)$ of Section 1) has already been determined

(Section 4). Suppose that N functions $x^1(s), \cdots, x^N(s)$, where $N \geqq 1$, have been determined such that $x^k(s)$ is a solution of (4′) satisfying

$$x^1(s) = x_1(s) \tag{38}$$

or, in case $2 \leqq k \leqq N$, the equations

$$x^k(s) = sx^{k-1}(s) + x_k(s) \tag{39}$$

and

$$d^l x_k / ds^l = \sum_{h=0}^{k-2} s^h w_{hl}, \qquad l = 0, 1, \cdots, n-1, \tag{40}$$

where the functions w_{hl} are uniformly almost periodic for $\sigma \geqq 0$ and of type (2′) and (3). The existence of an x^N, in the case $N = 1$, has been proved in Section 4. In line with the assertions represented by (6′), define the function $x^{N+1}(s)$, in case $n \geqq N + 1$, by

$$x^{N+1}(s) = sx^N(s) + x_{N+1}(s). \tag{41}$$

(Actually, if $n = 1$, so that $n \geqq N + 1$ cannot hold, the proof of Theorem I′ is already complete.) As is readily verified, in order that $x^{N+1}(s)$ be a solution of (4′), $x_{N+1}(s)$ must satisfy the equation,

$$\sum_{l=0}^{n} F_l(s) d^l x_{N+1}/ds^l = g(s), \tag{42}$$

where

$$g(s) = -\sum_{l=1}^{n} lF_l(s) d^{l-1}x^N/ds^{l-1}.$$

Hence, properties (2′) and (3) of the functions (1′), together with the properties of the functions $x^1, \cdots, x^N$, as given by (38), (39) and (40), imply that the present function $g(s)$, defined above, may be identified with that defined by (26), where $M = N - 1$. The only conceivable trouble might arise from the term $nd^{n-1}x^N/ds^{n-1}$ of the last summation. However, since $n - 1 \geqq N$, it follows from (38)-(40) that this function alone is of type (26) as are also the remaining terms (hence the whole) of the summation. Thus, as was shown in Section 5, (42) possesses a solution x_{N+1}, where

$$d^l x_{N+1}/ds^l = \sum_{h=0}^{N-1} s^h z_{hl}(s), \qquad l = 0, 1, \cdots, n-1, \tag{43}$$

and the functions z_{hl} are uniformly almost periodic for $\sigma \geqq 0$ and of type (2′) and (3). It follows from (38), (41) and (43) that (38), (39) and (40) remain valid if N is replaced by $N + 1$. It is clear that if the functions $x^k(s)$ are defined by (41) for $k = 2, 3, \cdots, N$ (the function x^1 being that determined in Section 4), then one obtains n solutions $x^1, \cdots, x^n$ of (4′) possessing the structures (6′), (7′) described in Theorem I′; cf. Section 2. That these n functions are independent is clear. In fact, if s is real, it is easily verified that $x^k(s)/s^{k-1} \to 1$ as $s \to \infty$. In particular, $x^1(s)$ is the only bounded solution of (4′). The assertions regarding the exponents of the functions $x_{kh}(s)$ of (7′) are clear from the proof.

PART II.

7. Results analogous to those obtained in Part I holds in case the functions (1) are absolutely convergent Laplace integrals. In fact, the following theorems will be proved:

THEOREM II. *On the real half line* $0 \leqq t < \infty$, *let each of the functions* (1) *be a Laplace integral of the type*

$$(44) \qquad f_k(t) = \int_1^\infty e^{-tu} d\alpha_k(u), \qquad 0 \leqq t < \infty, k = 0, 1, \cdots, n-1,$$

where

$$(45) \qquad [\alpha_k] = \int_1^\infty | d\alpha_k(u) | < \infty, \qquad k = 0, 1, \cdots, n-1,$$

and define $f_n(t)$ *by* (5). *Then* (4) *possesses n linearly independent solutions* $x = x^1(t), \cdots, x^n(t)$ *which can be chosen so as to possess the recursive structure* (6), *where* $x^1(t) = x_1(t)$ *is defined by*

$$(46) \qquad x^1(t) = 1 + \int_1^\infty e^{-tu} d\beta_1(u),$$

the remaining functions $x_k(t)$ *have the structure*

$$(47) \qquad x_k(t) = \sum_{h=0}^{k-2} t^h \int_1^\infty e^{-tu} d\beta_{kh}(u), \qquad k = 2, 3, \cdots, n,$$

and the functions β_1, β_{kh} *satisfy the relations*

$$(48) \qquad \int_1^\infty u^n \, | d\beta_1(u) | < \infty, \qquad \int_1^\infty | d\beta_{kh}(u) | < \infty;$$

finally, the spectra of the functions β_1 *and* β_{kh} *are contained in the closure* *of the set of those t-values representable as* $n_1t_1 + \cdots + n_mt_m$, *where* $m, n_1, \cdots, n_m$ *are positive integers and* $t_1, \cdots, t_m$ *are points belonging to the spectra of the functions* $a_0, a_1, \cdots, a_{n-1}$; cf. [4], p. 443.

Similar results hold for the homogeneous equation (8) if g is of the same type as the functions, f_k, $k = 0, 1, \cdots, n-1$.

The proof of Theorem II proceeds along the same lines as that of Theorem I′, the multiplication of two uniformly almost periodic functions in the latter case being now replaced by the convolution of two Laplace integrals. The rôle played by the above mentioned multiplication, in being responsible for the successive increases by one of the lower bound for the exponents involved in the successive approximations, is now taken over by the "additive rule" for spectra in the case of convolutions; cf. [3], p. 335.

8. *Proof of Theorem* II. The existence of the function $x^1(t)$ will first be shown. It is easily verified that the function

$$y_1(t) = \int_1^\infty e^{-tu} d\gamma_1(u), \tag{49}$$

where $\gamma_1(u)$ is defined by

$$(-u)^n d\gamma_1(u) = -da_0(u), \tag{50}$$

satisfies, for $0 \leqq t < \infty$, the differential equation

$$d^n y_{m+1}/dt^n = -\sum_{l=0}^{n-1} f_l(t)\, d^l y_m/dt^l, \qquad y_0(t) \equiv 1, \tag{51}$$

in which $m = 0$, so that

$$d^n y_1/dt^n = -f_0(t), \qquad 0 \leqq t < \infty. \tag{52}$$

Suppose that $y_m(t)$, where $m \geqq 1$, has been determined so that

$$y_m(t) = \int_1^\infty e^{-tu} d\gamma_m(u), \tag{53}$$

where the spectrum of $\gamma_m(t)$ does not extend below m, and where (cf. (45))

$$\delta_m = \int_1^\infty u^{n-1} \,|\, d\gamma_m(u) \,|\, < \infty. \tag{54}$$

That (53) and (54) hold for $m = 1$ follows from (49), (50), and (45). By (51) and (54),

$$\sum_{l=0}^{n-1} f_l(t)\, d^l y_m/dt^l = \sum_{l=0}^{n-1} \int_1^\infty e^{-tu} d(a_l * a_{ml})(u),$$

where

$$da_{ml}(u) = (-u)^l d\gamma_m(u), \qquad l = 0, 1, \cdots, n-1,$$

and the asterisk is the symbol of convolution. If $\gamma_{m+1}(u)$ is any function satisfying

$$(-u)^n d\gamma_{m+1}(u) = -d \sum_{l=0}^{n-1} a_l * a_{lm},$$

(so that, except for a normalization constant $\gamma_{m+1}(1)$, any function $\gamma_{m+1}(u)$ is uniquely defined at its continuity points, only $\gamma_{m+1}(u-0)$ and $\gamma_{m+1}(u+0)$ being relevant; cf. [3], p. 334), it is seen that the function $y_{m+1}(s)$ defined by (53), if m is replaced by $m+1$, satisfies (51) for $0 \leqq t < \infty$ and, since the spectrum of $\gamma_{m+1}(u)$ does not extend below $m+1$, that

$$\delta_{m+1} = \int_0^\infty u^{n-1} \mid d\gamma_{m+1}(u) \mid \leqq nC\delta_m/(m+1),$$

where $C = \max([a_0], [a_1], \cdots, [a_{n-1}])$. (Use is made here of the convolution inequalities $[a_l * a_{lm}] \leqq [a_l][a_{lm}]$; cf. [4], p. 446.) Thus

$$(55) \qquad \delta_m = O(a^m/m!), \qquad m = 1, 2, \cdots,$$

where a denotes some positive constant independent of m (cf. [4], p. 446). If $\beta_1(u)$ is defined by

$$(56) \qquad \beta_1(u) = \sum_{m=1}^\infty \gamma_m(u),$$

it is clear that the spectrum of $\beta_1(t)$ has the properties described in Section 7, and that the function $x^1(t)$ defined by (46) is a non-trivial solution of (4) (in fact, $x^1(t) \to 1$ as $t \to \infty$). That

$$(57) \qquad \int_1^\infty u^{n-1} \mid d\beta_1(u) \mid < \infty,$$

follows from (55) and (56). If $t > 0$, it is seen from (4) and (46) that

$$\int_1^\infty (-u)^n e^{-tu} d\beta_1(u) = -\sum_{l=0}^{n-1} \int_1^\infty e^{-tu} d(a_l * \beta_1{}^l),$$

where $\beta_1^l = (-u)^l d\beta_1$, $l = 0, 1, \cdots, n-1$. Hence

$$(-u)^n d\beta_1(w) = -\sum_{l=0}^{u-1} d(\alpha_l * \beta_1^l)$$

and relations (45) and (57) imply, via (57) and the relation $[\alpha_l * \beta_1^l] \leqq [\alpha_l][\beta_1^l]$, the first inequality of (48).

The remainder of the proof will hinge upon the following lemma, corresponding to Lemma 2 of Section 3:

LEMMA 3. *On the half-line* $0 \leqq t < \infty$, *let* $h_0(t), h_1(t), \cdots, h_M(t)$ *denote* $M+1$ *Laplace integrals of the type*

$$h_k(t) = \int_v^\infty e^{-tu} d\sigma_k(u), \qquad v > 0, n = 0, 1, \cdots, M,$$

where $\mu = \max([\sigma_0], [\sigma_1], \cdots, [\sigma_M]) < \infty$. *Then the differential equation*

$$d^n x/dt^n = \sum_{k=0}^{M} t^k h_k(t), \qquad 0 \leqq t < \infty,$$

possesses, for $0 \leqq t < \infty$, *a solution such that*

$$d^l x/dt^l = \sum_{k=0}^{M} t^k x_{kl}, \qquad l = 0, 1, \cdots, n-1,$$

and the functions x_{kl} *are such that*

$$x_{kl}(t) = \int_v^\infty e^{-tu} d\alpha_{kl}(w), \qquad k = 0, 1, \cdots, M \text{ and } l = 0, 1, \cdots, n-1,$$

and

$$[\alpha_{kl}] \leqq C_M \mu v^{-1}, \qquad k = 0, 1, \cdots, M \text{ and } l = 0, 1, \cdots, n-1$$

where C_M *denotes a constant depending only on* M.

The proof of Lemma 3 is similar to that of Lemma 2 and can therefore be omitted.

The remainder of the proof of Theorem II will now be sketched. By use of Lemma 3 and a method similar to that used in Section 5, one shows that the inhomogeneous differential equation

$$\sum_{h=0}^{n} f_h(t) d^h x/dt^h = \sum_{k=0}^{M} t^k h_k, \qquad 0 \leqq t < \infty,$$

where the functions $h_k(t)$ are Laplace integrals of the type (44) and (45), has a solution $x(t)$ such that

$$d^lx/dt^l = \sum_{k=0}^{M} t^k z_{kl}, \qquad l = 0, 1, \cdots, n-1,$$

and the functions z_{kl} are Laplace integrals of the type (44) and (45). The application of the above results in that portion of the proof of Theorem II, corresponding to Section 6 for Theorem I′, is clear.

That Theorem II can be generalized to the complex case by replacing the real variable t, $0 \leqq t < \infty$, by $s = \sigma + it$, where $\sigma \geqq 0$ and $-\infty < t < \infty$, is seen from the remark of [4], p. 447.

THE INSTITUTE FOR ADVANCED STUDY
AND
THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] A. S. Besicovitch, *Almost Periodic Function*, Cambridge (1932).

[2] P. Hartman and A. Wintner, "On the classical existence theorem of linear differential equations," *American Journal of Mathematics*, vol. 71 (1949), pp. 859-864.

[3] A. Wintner, "On the classical existence theorem of linear differential equations," *ibid.*, vol. 71 (1949), pp. 331-338.

[4] ———, "On the existence of Laplace solutions for linear differential equations of second order," *ibid.*, vol. 72 (1950), pp. 442-450.

GENERALIZATION OF A THEOREM OF MANDELBROJT.*

By JAMES A. JENKINS.

1. Some years ago S. Mandelbrojt [6] initiated the study of quasi-analytic classes of functions integrable over a finite interval. This work was continued and extended by Levin and Lifschitz [5] who considered the relationship between the admissible integral orders of zeros of such a function and the gaps occurring in its Fourier series. Let $f(t)$ be a function summable over the interval $(-\pi, \pi)$ and have the associated Fourier series

$$f(t) \sim \tfrac{1}{2}a_0 + \sum_{n=1}^{\infty} (a_n \cos nt + b_n \sin nt).$$

Let $\psi(\alpha)$ be a continuous, non-negative, non-decreasing function defined for $0 \leqq \alpha \leqq 2\pi$. Then we say that $f(t)$ has a left-hand zero of integral order $\psi(\alpha)$ at $t = \pi$ if

$$\phi(\alpha) \equiv \int_{\pi-\alpha}^{\pi} |f(t)|\, dt \leqq \psi(\alpha) \qquad (0 \leqq \alpha \leqq 2\pi).$$

Alternatively one could consider right-hand zeros of this type or such zeros situated at other points. In terms of a function derived from $\psi(\alpha)$ and a function measuring the density of the indices of non-vanishing terms of the Fourier series, Levin and Lifschitz gave a condition implying that $f(t)$ vanishes almost everywhere on the interval $(-\pi, \pi)$. I. I. Hirschman and the author gave a simple proof of a refinement of this result [2].

The question arises whether something of a similar nature can be done for functions defined on the infinite interval $(-\infty, \infty)$ and it seems there is a rather natural connection with the general principle of Wiener that a function and its Fourier transform cannot both be too small at infinity. We may suppose that $f(t)$ and $\phi(x)$ are in $L^2(-\infty, \infty)$ and are transforms of one another in the sense of Plancherel's Theorem. G. H. Hardy [1] and G. W. Morgan [7] proved results of this type where bounds were given for the modulus of each function for $|t|$ and $|x|$ large. Ingham [4] gave a related result, assuming that f vanished outside a finite interval and that an integral related to ϕ diverged.

---

* Received January 13, 1951.

Recently I. I. Hirschman [3] proved a result of this type, giving a bound for the modulus of f as $t \to +\infty$ and assuming a certain rate of divergence of an integral associated with ϕ. This result is significant chiefly when f tends to zero very rapidly as $t \to \infty$ and ϕ tends to zero rather slowly. Thus it appears that his result is most closely related to Ingham's and indeed does not include the results of Hardy and Morgan in any form.

The object of the present paper is to give a generalization of the preceding results. First of all the bounds for the modulus are replaced by bounds on certain integrals. Secondly the rather special assumptions placed on ϕ by Hirschman are removed so that the result obtained includes the main part of the results of Hardy and Morgan and a more general result than Ingham's in his special case. Indeed when f vanishes say for t greater than a certain value our result reduces to one direction of the fundamental result of Paley and Wiener [8], p. 16. The principal method used, as in [2] and [3], is the principle of subordination for the logarithm of the modulus of an analytic function.

2. THEOREM 1. *Let $f(t)$ and $\phi(x)$ be in $L^2(-\infty, \infty)$ and be transforms of each other in the sense of Plancherel's Theorem, i. e.,*

$$\phi(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-ixt} f(t)\, dt, \qquad f(t) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{itx} \phi(x)\, dx,$$

where equality is to be understood in the sense of the mean square limit. Let

$$U(r) = \pi^{-1} \int_{-r}^{r} (1+x^2)^{-1} \log |\phi(x)|\, dx \tag{1}$$

tend to $-\infty$ as $r \to \infty$. Let there exist a function $V(y)$, non-decreasing for $y > 0$ such that

$$\int_{V(y)}^{\infty} e^{ty} |f(t)|\, dt = O(e^{yV(y)}) \text{ as } y \to \infty. \tag{2}$$

Then if

$$\liminf_{r \to \infty} (V(r) + U(r)) = -\infty, \tag{3}$$

f and ϕ are both almost everywhere zero.

We observe first that the condition (2) implies that

$$F(z) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-itz} f(t)\, dt \qquad (z = x + iy)$$

represents a function regular for $y > 0$. It is easily deduced that

$$\lim_{y \to 0+} \| F(x+iy) - \phi(x) \|_2 = 0, \tag{4}$$

where $\| g(x) \|_2$ denotes the L^2 norm $\left(\int_{-\infty}^{\infty} | g(x) |^2 dx \right)^{\frac{1}{2}}$ of a function $g(x)$ in $L^2(-\infty, \infty)$. Moreover, for $y > 0$

$$\begin{aligned} -\int_{-\infty}^{\infty} e^{-itz} f(t)\, dt &\leqq \int_{-\infty}^{V(y)} e^{ty} | f(t) |\, dt + \int_{V(y)}^{\infty} e^{ty} | f(t) |\, dt \\ &\leqq \left(\int_{-\infty}^{V(y)} e^{2ty} dt \right)^{\frac{1}{2}} \left(\int_{-\infty}^{V(y)} | f(t) |^2\, dt \right)^{\frac{1}{2}} + O(e^{yV(y)}). \end{aligned}$$

Hence $| F(z) | \leqq A y^{-\frac{1}{2}} e^{yV(y)}$, $y < 1$ and $| F(z) | \leqq A e^{yV(y)}$, $y \geqq 1$ for an appropriate constant A.

Now let us consider a circle, centre the origin, of radius $r > 1$. Let it cut off on the line $y = \epsilon (0 < \epsilon < 1)$ a chord $S(\epsilon)$. Above the chord lies an arc $C(\epsilon)$ of the circle. Together $S(\epsilon)$ and $C(\epsilon)$ bound a domain $D(\epsilon)$ which contains the point $z = i$ and in which $F(z)$ is bounded. (The bound may depend on ϵ.) Hence we can use the principle of majorization:

$$\begin{aligned} \log | F(i) | \leqq & \int_{S(\epsilon)} K_\epsilon(x, \epsilon) \log | F(x + i\epsilon) |\, dx \\ & + \int_{C(\epsilon)} K_\epsilon(x, y) \log | F(x + iy) |\, ds \end{aligned}$$

where $K_\epsilon(x, y)$ is the kernel giving the value of a harmonic function at the point $z = i$ in terms of its values on the boundary of $D(\epsilon)$.

In the customary notation, we write

$$\log | F(x + i\epsilon) | = \log^+ | F(x + i\epsilon) | + \log^- | F(x + i\epsilon) |.$$

As $\epsilon \to 0$, $\log^+ | F(x + i\epsilon) |$ tends in the L^2 norm to $\log^+ | \phi(x) |$, each of these clearly belonging to $L^2(-\infty, \infty)$. Thus the corresponding part of the integral over $S(\epsilon)$ tends to a finite quantity which has a uniform bound for all r. Moreover by the principle of extension of domain, replacing the circular segment $D(\epsilon)$ by a half plane bounded by $y = \epsilon$, we have

$$\begin{aligned} & \int_{S(\epsilon)} \log^- | F(x + i\epsilon) | K_\epsilon(x, \epsilon)\, dx \\ & \qquad \leqq \int_{S(\epsilon)} \log^- | F(x + i\epsilon) | \pi^{-1} (1 - \epsilon) (x^2 + (1 - \epsilon)^2)^{-1} dx. \end{aligned}$$

Using the same argument as is used to prove Fatou's Theorem [9], p. 346, we find

$$\limsup_{\epsilon\to 0+}\int_{S(\epsilon)}\log^-|F(x+i\epsilon)|\,\pi^{-1}(1-\epsilon)(x^2+(1-\epsilon)^2)^{-1}dx$$

$$\leqq \pi^{-1}\int_{-r}^{r}(1+x^2)^{-1}\log^-|\phi(x)|\,dx.$$

On the other hand, for the integral over $C(\epsilon)$ we regard the two bounds for $|F(z)|$ according as $y<1$ *or* $y\geqq 1$ and obtain the majorant expression

$$\int_{C(\epsilon)}((A+yV(y))K_\epsilon(x,y)\,ds-\int_{C(\epsilon)-C(1)}\tfrac{1}{2}(\log y)K_\epsilon(x,y)\,ds.$$

Since $V(y)$ is increasing we can, in the first integral, replace $V(y)$ by $V(r)$ and obtain a still larger result. The new expression then tends, as $\epsilon\to 0$, to $V(r)$ plus a constant independent of r. The second integral converges to a quantity readily seen to be $o(r)$. Thus we have $\log|F(i)|\leqq V(r)+U(r)+O(1)$. Hence the condition (3) implies $F(i)=0$. If we did not have $F(z)\equiv 0$ there would be an integer m such that $F(z)$ had a zero of exactly order m at $z=i$. However the same argument applied to $F(z)(z-i)^{-m}=\Phi(z)$ would imply, since $\log|z-i|$ is uniformly bounded below on our contours as $\epsilon\to 0$, that $\Phi(i)=0$. Hence $F(z)\equiv 0$ from which the conclusion of our theorem follows.

We observe in addition that for $\phi(x)$ to vanish over an interval or to behave otherwise so that the integral in (1) diverges to $-\infty$ when taken for finite r already implies that f and ϕ are both almost everywhere zero.

3. In the paper cited, Hirschman applied his result to certain classes of quasi-analytic functions. In order to do this it was necessary to restrict consideration to those classes which contain the analytic class. We will now show that this restriction is susbtantially inessential.

First we recall a few definitions concerning quasi-analytic functions. Let $C\{M_n\}$ denote the class of functions $g(t)$, defined and infinitely differentiable for $-\infty<t<\infty$ and such that

$$|g^{(n)}(t)|\leqq Ak^nM_n \qquad (-\infty<t<\infty\,;\,n=0,1,2,\cdots)$$

with A, k constants which may depend on $g(t)$. The convex regularization $\{M_n{}^c\}$ of the sequence $\{M_n\}$ is given by

$$T(r)=\operatorname*{Max}_{n\geqq 0}(r^n/M_n),\qquad M_n{}^c=\operatorname*{Max}_{r\geqq 0}(r^n/T(r)).$$

THEOREM 2. *Let* $g(t)$ *belong to the class* $C\{M_n\}$ *and satisfy*

$$\int_{V(y)}^{\infty} e^{ty} \, | \, g(t) \, | \, dt = O(e^{yV(y)}), \qquad y \to \infty,$$

with $V(y)$ *non-decreasing for* $y > 0$. *Let*

$$W(r,h) = (2/\pi) \int^{r} (1 + x^2)^{-1} \log T(x/h) \, dx$$

satisfy for some $h > k$

$$\liminf_{r \to \infty} \, (V(r) - W(r,h)) = -\infty. \tag{5}$$

Finally, let $\lim\limits_{n \to \infty} (M_n^c)^{1/n} = \infty$. *Then* $g(t) \equiv 0$.

Indeed, let $f(t) = g(t)(\sin t)/t$ and let $\phi(x)$ be the transform of $f(t)$ in the sense of Parseval's Theorem. Then Hirschman showed that, under the given hypotheses, for every $k' > k$, $T(|\,x\,|/k')\phi(x)$ is in L^2 $(-\infty, \infty)$. Let us denote this function by $\psi(t)$ and take k' to have the value h in (5) above. Now

$$\int_{-r}^{r} (1 + x^2)^{-1} \log |\,\phi(x)\,| \, dx$$
$$= \int_{-r}^{r} (1 + x^2)^{-1} \log |\,\psi(x)\,| \, dx - \int_{-r}^{r} (1 + x^2)^{-1} \log T(|\,x\,|/h) \, dx$$

and we easily verify that $\int_{-r}^{r} (1 + x^2)^{-1} \log |\,\psi(x)\,| \, dx < M$, independently of r. Thus

$$\int_{-r}^{r} (1 + x^2)^{-1} \log |\,\phi(x)\,| \, dx < - \int_{-r}^{r} (1 + x^2)^{-1} \log T(|\,x\,|/h) \, dx + M.$$

Clearly $f(t)$ satisfies the condition (2) in Theorem 1 and the assumption (5) of the present theorem implies condition (3) of Theorem 1. Hence both $f(t)$ and $g(t)$ are identically zero since continuous.

We observe that in addition to the generalization indicated above we have replaced the assumption on the asymptotic behavior of $g(t)$ used by Hirschman by an average assumption.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] G. H. Hardy, " A theorem concerning Fourier transforms," *Journal of the London Mathematical Society*, vol. 8 (1933), pp. 227-231.

[2] I. I. Hirschman, Jr. and J. A. Jenkins, " Note on a result of Levin and Lifschitz," *Proceedings of the American Mathematical Society*, vol. 1 (1950), pp. 390-393.

[3] I. I. Hirschman, Jr., " On the behavior of Fourier transforms at infinity and on quasi-analytic classes of functions," *American Journal of Mathematics*, vol. 72 (1950), pp. 200-213.

[4] A. E. Ingham, "A note on Fourier transforms," *Journal of the London Mathematical Society*, vol. 9 (1934), pp. 29-32.

[5] B. Levin and M. Lifschitz, " Quasi-analytic functions represented by Fourier series," *Matematicheskii Sbornik*, vol. 9 (1941), pp. 693-711.

[6] S. Mandelbrojt, *Séries de Fourier et classes quasi-analytiques de fonctions*, Paris, 1935.

[7] G. W. Morgan, "A note on Fourier transforms," *Journal of the London Mathematical Society*, vol. 9 (1934), pp. 187-192.

[8] R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, New York, 1934.

[9] E. C. Titchmarsh, *The theory of functions*, Oxford, 1932.

THE THEOREM OF RIEMANN-ROCH ON COMPACT ANALYTIC SURFACES.*

By Kunihiko Kodaira.

The main purpose of the present paper is to prove the theorem of Riemann-Roch on compact complex analytic manifolds of complex dimension 2 with Kählerian metrics[1] by means of the theory of harmonic integrals. Suppose such a manifold $\mathfrak{M}$ as given and denote for an arbitrary divisor D on $\mathfrak{M}$ the space consisting of all meromorphic functions F on $\mathfrak{M}$ which are multiples of $-D$ by $\mathfrak{F}(D)$. The problem concerning the theorem of Riemann-Roch consists in expressing the dimension of the linear space $\mathfrak{F}(D)$ in terms of the "characters" of D such as the topological intersection-number of D with itself, the virtual genus of D, etc.[2] In this paper, we shall mainly consider the case that D is an arbitrary (reducible or irreducible) curve Γ. In Section **1**, we shall first summarize the theory of harmonic currents due to G. de Rham[3] and several results of W. V. D. Hodge[4] concerning analytic manifolds with Kählerian metrics; then we shall prove a theorem concerning the existence of analytic differentials with given singularities. In Section **2**, we shall prove a theorem due to A. Weil and, by means of that theorem, introduce the notion of the characteristic divisor class $\{\mathfrak{d}\}$ on Γ which corresponds to the characteristic series of Italian geometers. In Section 3, we first introduce the linear space $\mathfrak{f}(\Gamma, \mathfrak{d})$ consisting of meromorphic functions on Γ associated with an arbitrarily fixed characteristic divisor $\mathfrak{d} \varepsilon \{\mathfrak{d}\}$ and show that the singularities of any meromorphic function $F \varepsilon \mathfrak{F}(\Gamma)$ can be represented by a meromorphic function $f \varepsilon \mathfrak{f}(\Gamma, \mathfrak{d})$; secondly, using the existence theorem in Section **1**, we prove that, for every $f \varepsilon \mathfrak{f}(\Gamma, \mathfrak{d})$, there exists on $\mathfrak{M}$ at least one additive meromorphic function which is a multiple of $-\Gamma$ and has the singularities represented by f on Γ; thirdly, calculating the period of this additive meromorphic function, we find a necessary and sufficient condition for $f \varepsilon \mathfrak{f}(\Gamma, \mathfrak{d})$ in order that f should represent the singularities of a meromorphic function $F \varepsilon \mathfrak{F}(\Gamma)$ and deduce an expression of

---

* Received December 20, 1950.

[1] Kähler [13].

[2] Zariski [30], p. 66.

[3] de Rham and Kodaira [20], Part I.

[4] Hodge [12].

the difference $\dim \mathfrak{f}(\Gamma, \mathfrak{d}) - \dim \mathfrak{F}(\Gamma)$ in terms of the simple differentials of the first kind on $\mathfrak{M}$. Our method applied here is a natural generalization of that of H. Weyl.[5] The dimension of $\mathfrak{f}(\Gamma, \mathfrak{d})$ is given by the theorem of Riemann-Roch on reducible curves which will be proved in Section **4**. Combining the result of Section **3** with that of Section **4**, we shall prove, in the following Section **5**, a formula expressing $\dim \mathfrak{F}(\Gamma)$ in terms of the topological intersection number of Γ with itself, the virtual genus of Γ and two non-negative integers corresponding to the index of speciality appearing in the theorem of Riemann-Roch on algebraic curves. This formula may be called the *theorem of Riemann-Roch* on compact analytic surfaces with Kählerian metrics, since it contains the theorem of Riemann-Roch for irreducible linear systems on algebraic surfaces.[6] In the last part of this Section, we shall deduce a theorem of Castelnuovo[7] concerning the superabundance of irreducible linear systems. In the following Section **6**, we first prove a formula expressing the number of linearly independent double differentials on $\mathfrak{M}$ which are multiples of $-\Gamma$ in terms of the virtual genus of Γ, the number of connected components of Γ and the number of simple Picard integrals of the first kind which are constant on each connected components of Γ, and then we deduce from that formula a theorem of F. Enriques[8] concerning the deficiency of the series cut out on a generic curve of an arbitrary irreducible linear system by its adjoint system and a theorem of F. Severi[9] concerning the superabundance of the adjoint systems of irreducible linear systems. Applying our results to algebraic surfaces, we shall prove, in the final Section **7**, several classical results of Italian geometers, e. g. a theorem of G. Castelnuovo[10] concerning the deficiency of characteristic series. A strict algebraic proof of the theorem of Riemann-Roch for arbitrary divisors on algebraic surfaces was given recently by O. Goldman.[11] In the last part of this Section, we shall show that the theorem of Riemann-Roch for arbitrary divisors on algebraic surfaces can be readily deduced from our results in the same manner as in the proof of O. Goldman.

The author wishes to express his grateful thanks to A. Weil for suggestions and encouragement given during the period of preparation of this paper. In particular, Theorem 2.1 and the substance of its proof, the substance of the proof of Theorem 3.2 and the notion of canonical cycles introduced in Section **5** are due to A. Weil. Again, the first part of Section **3** and the first part of Section **5** were revised according to suggestions of A. Weil. The author's best thanks are also due to O. Goldman and D. Gorenstein, who gave him the opportunity to study their manuscripts before publication at a very early stage of preparation of this paper. In addition to the method of deduction of the theorem of Riemann-Roch on algebraic surfaces in Section **7**, the author is indebted to O. Goldman in many details.

---

[5] Weyl [27], pp. 108-125.

[6] See Zariski [30], pp. 66-69.

[7] Castelnuovo [4]; see also Zariski [30], p. 71.

[8] Enriques [8]; see also Zariski [30],pp. 64-66.

[9] Severi [23]; cf. Zariski [30], p. 144.

[10] Castelnuovo [4]; see also Zariski [30], pp. 69-70.

[11] Goldman [9].

1. Currents on Kählerian manifolds. Let $\mathfrak{M}^{2n}$ be a compact analytic manifold of complex dimension n with a positive definite Kählerian metric[12]

$$ds^2 = 2 \sum_{\alpha,\beta=1}^{n} g_{\alpha\beta^*}(dz^\alpha d\bar{z}^\beta),$$

where $z = (z^1, z^2, \cdots, z^n)$ denotes the *local analytic coordinates* on $\mathfrak{M}^{2n}$. Such a manifold is called a *Kählerian manifold.* Putting

$$z^\alpha = x^{2\alpha-1} + ix^{2\alpha}, \qquad (\alpha = 1, 2, \cdots, n), \tag{1.1}$$

we introduce real coordinates $x^1, x^2, \cdots, x^{2n}$; then ds^2 can be written as

$$ds^2 = \sum_{j,k=1}^{2n} g_{jk}(dx^j dx^k), \qquad (g_{jk} = g_{kj}).$$

Thus $\mathfrak{M}^{2n}$ is a $2n$-dimensional Riemann manifold with a positive definite metric ds^2. Moreover $\mathfrak{M}^{2n}$ is *orientable*. We choose the orientation of $\mathfrak{M}^{2n}$ so that the system of local coordinates introduced in (1.1) is *positive* with respect to this orientation. By a *p-form* on $\mathfrak{M}^{2n}$ we shall mean an exterior differential form

$$\Phi = \Phi(x) = (1/p!) \sum \Phi_{jk\ldots l}(x) dx^j dx^k \cdots dx^l$$

of rank p with *complex* coefficients $\Phi_{jk\ldots l}$ defined in a domain $\mathfrak{D} \subseteqq \mathfrak{M}^{2n}$. Φ is said to be of class C^ν, if each coefficient $\Phi_{jk\ldots l}$ is a function of class C^ν of $x^1, x^2, \cdots, x^{2n}$. We denote by $d\Phi$ the *exterior derivative* of Φ and by $*\Phi$ the *adjoint form* of Φ; again the *exterior product* of two forms Φ and Ψ will be denoted by $\Phi \cdot \Psi$. The closure of the subset $\{x \mid \Phi(x) \neq 0\}$ will be called the *carrier* of Φ and be denoted by $|\Phi|$.

Now we shall introduce the notion of currents.[13] Let $\mathfrak{D} \subseteqq \mathfrak{M}^{2n}$ be an arbitrary fixed domain and $\{\Phi\}$ be the linear space consisting of all $(2n-p)$-

---

[12] We insert here a parenthesis to designate that the product of dz^α and $d\bar{z}^\beta$ is ordinary product.

[13] de Rham and Kodaira [20], Chap. II; Schwartz [21].

forms Φ of class C^∞ with $|\Phi| \subseteq \mathfrak{D}$. Then, a *p-current* in $\mathfrak{D}$ is, by definition, a linear functional $T[\Phi]$ defined on $\{\Phi\}$ which is *continuous* in the following sense: *For an arbitrary sequence* $\{\Phi^{(1)}, \Phi^{(2)}, \cdots, \Phi^{(m)}, \cdots\}$ *of forms* $\Phi^{(m)} \varepsilon \{\Phi\}$ *such that all* $|\Phi^{(m)}|$ *are contained in one and the same compact subset* $\mathfrak{R} \subseteq \mathfrak{D}$ *covered by a single system of local coordinates* $x^1, x^2, \cdots, x^{2n}$, *we have* $T[\Phi^{(m)}] \to 0 \, (m \to \infty)$ *if each partial derivative* $\partial^s \Phi^{(m)}{}_{jk\ldots l}/\partial x^p \partial x^q \cdots \partial x^r$ *converges uniformly to zero for* $m \to \infty$. We consider every continuous p-form Ψ defined in $\mathfrak{D}$ as a current by identifying Ψ with the linear functional

$$\Psi[\Phi] = \int_{\mathfrak{D}} \Psi \cdot \Phi; \tag{1.2}$$

similarly, we consider every differentiable $(2n-p)$-chain Γ as a p-current defined by $\Gamma[\Phi] = \int_\Gamma \Phi$. Thus *the notion of currents is a generalization of both notions of differential forms and chains.* We say that the p-current T vanishes at a point $\mathfrak{p} \varepsilon \mathfrak{D}$ and write $T = 0$ at $\mathfrak{p}$ if there exists a neighborhood $\mathfrak{U}_\mathfrak{p}$ of $\mathfrak{p}$ such that $T[\Phi] = 0$ for all Φ with $|\Phi| \subset \mathfrak{U}_\mathfrak{p}$. By the *carrier* $|T|$ of T will be meant the set consisting of all points $\mathfrak{p} \varepsilon \mathfrak{D}$ such that T does *not* vanish at $\mathfrak{p}$. We get from (1.2) the identities

$$d\Psi[\Phi] = (-1)^{p+1}\Psi[d\Phi] \text{ and } (*\Psi)[\Phi] = (-1)^p\Psi[*\Phi].$$

In view of these identities, we define the *exterior derivative* dT and the *adjoint* $*T$ of an arbitrary p-current T by

$$dT[\Phi] = (-1)^{p+1}T[d\Phi]; \qquad (*T)[\Phi] = (-1)^p T[*\Phi]. \tag{1.3}$$

Then we have $ddT = 0$ and $**T = (-1)^p T$. In case T is a $(2n-p)$-chain Γ, dT coincides with the boundary $\mathfrak{B}\Gamma$ of Γ up to the sign $(-1)^{p+1}$; $d\Gamma = (-1)^{p+1}\mathfrak{B}\Gamma$. We introduce further the operators δ and Δ defined respectively as $\delta = -*d*$, $\Delta = d\delta + \delta d$. Again, the *exterior product* $T \cdot \Psi$ of a p-current T and a q-form Ψ of class C^∞ is defined as $(T \cdot \Psi)[\Phi] = T[\Psi \cdot \Phi]$. For an arbitrary p-current T and p-form Φ of class C^∞ defined on the whole manifold $\mathfrak{M}^{2n}$, we define the *inner product* of T and Φ as $(T, \Phi) = (\Phi, T) = T[*\Phi]$. In case $T = \Psi$ is a p-form of class C^∞, we have

$$(\Psi, \Phi) = \int_{\mathfrak{M}} \Psi \cdot *\Phi = \int_{\mathfrak{M}} \Phi \cdot *\Psi;$$

especially $(\Phi, \bar{\Phi}) > 0$ if Φ does not vanish identically, where $\bar{\Phi}$ denotes the *conjugate* form $(1/p!) \sum \bar{\Phi}_{jk\ldots l} dx^j dx^k \cdots dx^l$. Again, we define the *conjugate* current $\bar{T}$ of T by the formula: $\bar{T}[\bar{\Phi}] = \overline{T[\Phi]}$. A p-current T is said to be regular at a point $\mathfrak{p} \varepsilon \mathfrak{D}$, if T coincides with a p-form of class C^∞ in some neighborhood $\mathfrak{U}_{\mathfrak{p}}$ of $\mathfrak{p}$; otherwise T is called *singular* at $\mathfrak{p}$. By the *singular set* of T will be meant the set of all singular points of T. The singular set of T will be denoted by $|T|_s$. Obviously $|T|_s$ is a subset of the carrier $|T|$. In case T is defined in the whole manifold $\mathfrak{M}^{2n}$, the singular set $|T|_s$ of T is compact and, for an arbitrary open set $\mathfrak{U} \supset |T|_s$, we can find a p-form Ψ of class C^∞ such that $|T - \Psi| \subset \mathfrak{U}$. Now, the *inner* product (S, T) of arbitrary p-currents, S, T in the whole manifold $\mathfrak{M}^{2n}$ can be defined, under the assumption that $|S|_s \cap |T|_s$ is empty, as follows: *Choosing p-forms* Φ *and* Ψ *of class* C^∞ *so that* $|T - \Psi| \cap |S - \Phi|$ *is empty, we put*

$$(S, T) = (\Phi, \Psi) + (\Phi, T - \Psi) + (S - \Phi, \Psi).$$

It can be easily verified that the value of (S, T) does not depend on the choice of Φ and Ψ. We have $(dS, T) = (S, \delta T)$, $(\Delta S, T) = (S, \Delta T)$; also $(*S, T) = (-1)^p (S, *T)$.

By a *harmonic p-form of the first kind* on $\mathfrak{M}^{2n}$ we shall mean a p-form e of class C^∞ satisfying $de = 0$ and $\delta e = 0$ everywhere on $\mathfrak{M}^{2n}$. *The number of linearly independent harmonic p-forms* of the first kind[14] is equal to the p-Betti number b^p of $\mathfrak{M}^{2n}$. Let $e_1, e_2, \cdots, e_b$ $(b = b^p)$ be linearly independent real harmonic p-forms of the first kind such that $(e_j, e_k) = \delta_{jk}$. By the help of these forms we associate with each p-current T the harmonic p-form $HT = \sum_k (T, e_k) e_k$, which will be called the *harmonic part* of T. The operator H is *self-adjoint* in the sense that $(HS, T) = (S, HT)$ for arbitrary S, T. Again H satisfies $*H = H*$.

THEOREM 1.1 (G. de Rham).[15] *There exists one (and only one) linear operator* G *mapping any p-current* T *(defined in the whole of* $\mathfrak{M}^{2n}$*) into a p-current* GT *which has the following properties:*

$$\Delta GT = G\Delta T = T - HT, \qquad HGT = GHT = 0.$$

This operator G satisfies $Gd = dG$, $G\delta = \delta G$; again G is *self-adjoint* in the sense that $(GS, T) = (S, GT)$ for arbitrary S, T such that $|S|_s \cap |T|_s$ is empty. The operators G and H are *real* in the sense that $\overline{GT} = G\bar{T}$, $\overline{HT} = H\bar{T}$.

THEOREM 1.2 (G. de Rham).[16] *Let* T *be a p-current in a domain*

---

[14] Hodge [12]. See also de Rham and Kodaira [20], Part I, de Rham and Bidal [19], de Rham [18], Kodaira [14], Chap. II.

[15] de Rham and Kodaira [20], pp. 63-66.

[16] de Rham and Kodaira [20], p. 61. This theorem is a generalization of H. Weyl's fundamental lemma in his method of orthogonal projection in potential theory. See Weyl [28]; cf. also Kodaira [14], Chap. III.

$\mathfrak{D} \subseteqq \mathfrak{M}^{2n}$. *If* ΔT *is regular at a point* $\mathfrak{p} \in \mathfrak{D}$, *then* T *is also regular at* $\mathfrak{p}$. *Again, if* dT *and* δT *are both regular at* $\mathfrak{p}$, *then* T *is also regular at* $\mathfrak{p}$.

By virtue of this theorem, it follows from the identity $\Delta GT = T - HT$ that, *if* T *is regular at* $\mathfrak{p}$, *then* GT *is also regular at* $\mathfrak{p}$. Inserting $T = \Delta GT + HT$ into (S, T), we get

$$(S, T) = (dS, GdT) + (\delta S, G\delta T) + (S, HT). \tag{1.4}$$

This formula enables us to generalize the notion of inner product (S, T) to the case that $|\, dS \,|_s \cap |\, dT \,|_s$ and $|\, \delta S \,|_s \cap |\, \delta T \,|_s$ are both empty. Let Γ, Z be arbitrary chains of dimensions $p, 2n - p$, respectively. Then $(^*\Gamma, Z)$ is defined by (1.4) if $|\, d\Gamma \,| \cap |\, Z \,|$ and $|\, dZ \,| \cap |\, \Gamma \,|$ are both empty; moreover, as was proved by G. de Rham,[17] $(^*\Gamma, Z)$ *coincides with the topological intersection number* $I(\Gamma, Z)$ of Γ and Z:

$$I(\Gamma, Z) = (^*\Gamma, Z).$$

Especially if Γ *and* Z *are both cycles, then*

$$I(\Gamma, Z) = (^*\Gamma, HZ) = \int_\Gamma HZ.$$

An arbitrary p-current T can be represented as a formal differential form

$$T = \sum_{r+s=p} (1/r\,!s\,!) \sum T_{\alpha_1 \ldots \alpha_r \beta^*_1 \ldots \beta^*_s} dz^{\alpha_1} \cdots dz^{\alpha_r} d\bar{z}^{\beta_1} \cdots d\bar{z}^{\beta_s},$$

where the coefficients $T_{\alpha_1 \ldots \alpha_r \beta^*_1 \ldots \beta^*_s}$ are *distributions* on the space of local coordinates $z^1, \cdots z^n$ in the sense of L. Schwartz.[18] By means of this expression we define operators C and Λ as follows:[19]

$$CT = \sum_{r+s=p} (1/r\,!s\,!) \sum (i)^{r-s} T_{\alpha_1 \ldots \alpha_r \beta^*_1 \ldots \beta^*_s} dz^{\alpha_1} \cdots dz^{\alpha_r} d\bar{z}^{\beta_1} \cdots d\bar{z}^{\beta_s},$$

$$\Lambda T = \sum_{r+s=p-2} (1/r\,!s\,!) \sum (-1)^r i g^{\alpha\beta^*} T_{\alpha\alpha_1 \ldots \alpha_r \beta^* \beta^*_1 \ldots \beta^*_s} dz^{\alpha_1} \cdots dz^{\alpha_r} d\bar{z}^{\beta_1} \cdots d\bar{z}^{\beta_s},$$

where $g^{\alpha\beta^*}$ are the quantities defined by $\sum_\gamma g^{\alpha\gamma^*} g_{\beta\gamma^*} = \delta^\alpha{}_\beta$. CT is a p-current; ΛT is a $(p-2)$-current (in case $p = 0$ or $p = 1$ we have to put $\Lambda T = 0$). *The operators* C *and* Λ *are commutative with* Δ, G *and* H.[20] Again we have

$$CCT = (-1)^p T, \tag{1.5}$$

$$CT[\Phi] = (-1)^p T[C\Phi] \tag{1.6}$$

---

[17] de Rham and Kodaira [20], p. 71.

[18] Schwartz [21].

[19] These two operators were introduced by W. V. D. Hodge. See Hodge [12], p. 171. The definition of C and Λ employed here are due to A. Weil. Cf. Weil [24].

[20] Hodge [12], pp. 165-168, Weil [24], p. 111, Eckmann and Guggenheimer [7].

for an arbitrary p-current T; furthermore we have the important formula[21]

$$\Lambda d - d\Lambda = C^{-1}\delta C. \tag{1.7}$$

A p-current T of the type $T = (1/r!t!)\sum T_{\alpha_1\ldots\alpha_r\beta^*_1\ldots\beta^*_t} dz^{\alpha_1}\cdots dz^{\alpha_r} d\bar{z}^{\beta_1}\cdots d\bar{z}^{\beta_t}$ is called a *pure p-current*[22] of type t. Obviously every p-current T can be decomposed uniquely into the sum $T = \sum_{t=0}^{p} \boldsymbol{P}_t T$ of the pure p-currents $\boldsymbol{P}_t T$ of types t $(t = 0, 1, \cdots, p)$. As was proved by B. Eckmann and H. Guggenheimer,[23] the "projection operators" $\boldsymbol{P}_t$ thus defined satisfy

$$\Delta \boldsymbol{P}_t = \boldsymbol{P}_t \Delta, \tag{1.8}$$

while we get readily

$$(T, \overline{\Phi}) = \sum_{t=0}^{p} (\boldsymbol{P}_t T, \overline{\boldsymbol{P}_t \Phi}), \tag{1.9}$$

where Φ is an arbitrary p-form of class C^∞. As one readily infers from (1.8) and (1.9), the complex normalized orthogonal bases $E_1, E_2, \cdots, E_b$ of harmonic p-forms of the first kind can be chosen so that each E_k is a *pure p-form.* Consequently we have

$$H\boldsymbol{P}_t = \boldsymbol{P}_t H \tag{1.10}$$

and therefore

$$G\boldsymbol{P}_t = \boldsymbol{P}_t G. \tag{1.11}$$

A p-form $\Phi = (1/p!)\sum \Phi_{\alpha_1\ldots\alpha_p} dz^{\alpha_1}\cdots dz^{\alpha_p}$ of type 0 is said to be *holomorphic* (or *meromorphic*) at $\mathfrak{p}$, if in some neighborhood of $\mathfrak{p}$ the coefficients $\Phi_{\alpha_1\ldots\alpha_p}$ are holomorphic (or meromorphic) functions of $z^1, \cdots z^n$; if Φ is holomorphic (or meromorphic) at every point $\mathfrak{p}$ of a subset $\mathfrak{T} \subseteqq \mathfrak{M}^{2n}$, then we say that Φ is holomorphic (or meromorphic) on $\mathfrak{T}$. A meromorphic p-form Φ is said to be *regular* at $\mathfrak{p}$ if Φ is holomorphic at $\mathfrak{p}$; otherwise Φ is said to be *singular* at $\mathfrak{p}$. By a *p-ple differential* in a domain $\mathfrak{D} \subseteqq \mathfrak{M}^{2n}$ we shall mean a meromorphic p-form Φ defined in $\mathfrak{D}$ satisfying $d\Phi = 0$ in $\mathfrak{D}$ *except for the singular points of* Φ. A p-ple differential is called a *holomorphic p-ple differential* if it is a holomorphic p-form. Let Φ be an arbitrary holomorphic p-ple differential in $\mathfrak{D}$. Then since this Φ satisfies $C\Phi = i^p\Phi$ and $\Lambda\Phi = 0$, we get, using (1.7),

$$\delta\Phi = i^{-p}\delta C\Phi = i^{-p}C(\Lambda d - d\Lambda)\Phi = 0, \qquad \text{in } \mathfrak{D}.$$

---

[21] Weil [24], p. 111, Eckmann and Guggenheimer [7].

[22] See Hodge [12], pp. 188-192, Eckmann and Guggenheimer [7].

[23] Eckmann and Guggenheimer [7].

thus *an arbitrary holomorphic differential* Φ *satisfies* $\delta\Phi = 0$. A holomorphic p-ple differential defined in the whole of $\mathfrak{M}^{2n}$ is called a *p-ple differential of the first kind.* The above result shows that every p-ple differential of the first kind is a harmonic p-form of the first kind.

THEOREM 1.3. *Let T be a pure p-current of type 0 in a domain $\mathfrak{D} \subseteqq \mathfrak{M}^{2n}$. If T satisfies $dT = 0$ in $\mathfrak{D}$, then T is a holomorphic p-ple differential in $\mathfrak{D}$.*

Proof. Since T satisfies $CT = i^pT$ and $\Lambda T = 0$, we get

$$\delta T = i^{-p}\delta CT = i^{-p}C(\Lambda d - d\Lambda)T = 0, \text{ in } \mathfrak{D}.$$

Hence, by virtue of Theorem 1.2, T is a p-form of class C^∞; however, since T is of type 0, T is represented as $T = (1/p!) \sum \Phi_{\alpha_1\alpha_2\dots\alpha_p} dz^{\alpha_1} \cdots dz^{\alpha_p}$. Now the relation $dT = 0$ implies that $\partial\Phi_{\alpha_1\alpha_2\dots\alpha_p}/\partial\bar{z}^\beta = 0$ $(\beta = 1, 2, \cdots, n)$, proving that each $\Phi_{\alpha_1\alpha_2\dots\alpha_p}$ is a holomorphic function of $z^1, z^2, \cdots, z^n$, q. e. d.

The integral $F(z) = \int^z \Phi$ of an arbitrary holomorphic simple differential Φ defined in $\mathfrak{D}$ is a many valued holomorphic function in $\mathfrak{D}$ and Φ coincides with its differential dF: $\Phi = dF$. The integral $\int^z \Phi$ of an arbitrary simple differential Φ defined *in the whole of* $\mathfrak{M}$ is called a *simple Picard integral.* The simple Picard integral $\int^z \Phi$ is said to be *of the first kind* if Φ is a differential of the first kind. In case Φ has any singularities, the simple Picard integral $\int^z \Phi$ is called of *the second* or *the third kind* according as it is locally single valued or it is not.[24] Let $A_1, A_2, \cdots, A_q$ be a base of the space of all simple differentials of the first kind. Then $A_1, A_2, \cdots, A_q$, *$\bar{A}_1, \bar{A}_2, \cdots, \bar{A}_q$ are linearly independent and constitute a base of the space of all harmonic 1-forms of the first kind.*[25] Thus the number q of linearly independent simple differentials of the first kind is equal to a half of the 1-Betti number b^1 of $\mathfrak{M}^{2n}$: $q = \frac{1}{2}b^1$. This number q will be called the *irregularity* of the manifold $\mathfrak{M}^{2n}$.

THEOREM 1.4. *Let T be a pure $(p+1)$-current of type 1 defined in the whole of $\mathfrak{M}^{2n}$ satisfying $dT = 0$ and $HT = 0$, p being a positive integer $\leqq n$. Then the p-current $\Theta = (d\Lambda G + i\delta G)T$ is a pure p-current of type 0 and satisfies $d\Theta = iT$; moreover Θ is a holomorphic p-ple differential in*

---

[24] See Zariski [30], p. 119.

[25] Hodge [12], pp. 191-192; Weil [24], pp. 111-112.

$\mathfrak{M}^{2n} - |T|$. *Let* Q *be an arbitrary* p*-current such that* $T = dQ$. *Then the integral of* Θ *over an arbitrary* p*-cycle* $Z \subset \mathfrak{M}^{2n} - |T|$ *is given by*

$$\int_Z \Theta = i(*Z, Q) + (-1)^{p-1} iQ[HZ]. \tag{1.12}$$

Proof. Since T is a pure $(p+1)$-current of type 1, GT is also a pure $(p+1)$-current of type 1, as (1.11) shows. Hence we get, using (1.7),

$$\delta GT = i^{1-p}\delta CGT = i^{1-p}C(\Lambda d - d\Lambda)GT = -i^{1-p}Cd\Lambda GT,$$

and therefore

$$\Theta = (1 + i^{-p}C)\, d\Lambda GT. \tag{1.13}$$

ΛGT is a pure $(p-1)$-current of type 0. Consequently we have

$$d\Lambda GT = \mathbf{P}_0 d\Lambda GT + \mathbf{P}_1 d\Lambda GT.$$

Combined with (1.13), this yields immediately $\Theta = 2\mathbf{P}_0 d\Lambda GT$, proving that Θ is a pure p-current of type 0. On the other hand, we have

$$d\Theta = iGd\delta T = iG\Delta T = iT - iHT,$$

proving that $d\Theta = iT$. This implies that $d\Theta = 0$ in $\mathfrak{M} - |T|$. Hence, by virtue of Theorem 1.3, Θ is a holomorphic differential in $\mathfrak{M} - |T|$. The integral $\int_Z \Theta = Z[\Theta]$ can be evaluated as follows: We have

$$Z[\Theta] = iZ[\delta GdQ] = (-1)^p i(Z, *\delta GdQ) = i(d*Z, GdQ),$$

while, by (1.4), $(*Z, Q) = (d*Z, GdQ) + (*Z, HQ)$. Hence we obtain

$$Z[\Theta] = i(*Z, Q) - i(*Z, HQ) = i(*Z, Q) + (-1)^{p-1} iQ[HZ],$$

proving (1.12).

2. Characteristic divisors.

Let $\mathfrak{M}$ be a compact Kählerian manifold of complex dimension 2. By a *curve* Γ on $\mathfrak{M}$ we shall mean a compact analytic curve in $\mathfrak{M}$, i. e. a compact analytic subvariety of $\mathfrak{M}$ of complex dimension 1. As is well known, Γ is decomposed into the sum: $\Gamma = \sum_{\nu=1}^{\mu} \Gamma_\nu$ of a finite number of *irreducible* curves Γ_ν. Each Γ_ν is a holomorphic image of the *non-singular model* $\mathbf{\Gamma}_\nu$ of Γ_ν: $\Gamma_\nu = \phi(\mathbf{\Gamma}_\nu)$. The direct sum $\mathbf{\Gamma} = \sum_{\nu=1}^{\mu} \mathbf{\Gamma}_\nu$ will be called the *non-singular model* of Γ. We denote points on $\mathbf{\Gamma}$ by $p, q, \cdots$, and points on $\mathfrak{M}$ by $\mathfrak{p}, \mathfrak{q}, \cdots$. The mapping $q \rightarrow \mathfrak{q} = \phi(q)$ from $\mathbf{\Gamma}$ onto Γ is one-to-one unless $\mathfrak{q}$ is a singular point of Γ. For a singular point $\mathfrak{q}$, the inverse image $\phi^{-1}(\mathfrak{q})$ consists of a finite number of points and each point $q \varepsilon \phi^{-1}(\mathfrak{q})$ correspond one-to-one to an irreducible branch Γ_q of Γ passing through $\mathfrak{q}$. Let τ_p be a local uniformization variable on $\mathbf{\Gamma}$ with the center p and $z = (z^1, z^2)$ be a system of local coordinates on $\mathfrak{M}$ with the origin $\mathfrak{p} = \phi(p)$. Then the mapping $q \rightarrow \mathfrak{q} = \phi(q)$ which maps the neighborhood U_p of p *onto* the branch $\mathbf{\Gamma}_p$ can be written as $z^1 = \phi_p{}^1(\tau_p)$, $z^2 = \phi_p{}^2(\tau_p)$, where $\phi_p{}^1$, $\phi_p{}^2$ are *holomorphic* functions of τ_p defined in U_p satisfying $\phi_p{}^1(0) = \phi_p{}^2(0) = 0$. On the other hand, Γ is represented at each point $\mathfrak{p} \varepsilon \mathfrak{M}$ by a *minimal local equation* $R_{\mathfrak{p}}(z^1, z^2) = 0$, where $R_{\mathfrak{p}}(z^1, z^2)$ is a holomorphic function of z^1, z^2 defined in a neighborhood $\mathfrak{U}_{\mathfrak{p}}$ of $\mathfrak{p}$. In case $\mathfrak{p} = \phi(p)$ is a simple point of Γ, we can choose z^1, z^2, τ_p so that $R_{\mathfrak{p}}(z^1, z^2) = z^1$ and that $z^2 = \tau_p$ on Γ_p. In what follows we write, for simplicity's sake, $R_p(z)$ for $R_{\phi(p)}(z)$. Again, in case it is not necessary to specify the center of the local uniformization variable τ_p, we write simply τ for τ_p and ϕ^α for $\phi_p{}^\alpha$ $(\alpha = 1, 2)$.

Now we choose for each $p \varepsilon \mathbf{\Gamma}$ a sufficiently small neighborhood U_p of p so that the closure $[\phi(U_p)]$ of $\phi(U_p)$ satisfies the following two conditions: i) $[\phi(U_p)] - \phi(p)$ contains no singular point of Γ ($\phi(p)$ may be a singular point of Γ) and ii) the local equation $R_p(z^1, z^2) = 0$ of Γ is useful at every point $\mathfrak{q}$ on $[\phi(U_p)]$. Consider two points p, q on $\mathbf{\Gamma}$ such that $U_p \cap U_q$ is not empty. Then the ratio $T_{pq}(z) = R_p(z)/R_q(z)$ is a non-vanishing holomorphic function of $z = (z^1, z^2)$ defined in some neighborhood of $[\phi(U_p \cap U_q)]$. Putting $t_{pq}(\tau) = T_{pq}(\phi^1(\tau), \phi^2(\tau))$, we get, for each ordered pair $\{p, q\}$ with $U_p \cap U_q \neq 0$, a non-vanishing holomorphic function $t_{pq} = t_{pq}(\tau)$ defined in $U_p \cap U_q$. It is clear that these functions t_{pq} satisfy the following relations:

$$(2.1) \qquad t_{pq}t_{qp} = 1, \qquad \text{in } U_p \cap U_q,$$

$$(2.2) \qquad t_{pq}t_{qr}t_{rp} = 1, \qquad \text{in } U_p \cap U_q \cap U_r.$$

THEOREM 2.1 (A. Weil). *We can find a system* $\{h_p | p \varepsilon \mathbf{\Gamma}\}$ *of meromorphic functions* h_p *defined respectively in* U_p *such that*

$$(2.3) \qquad h_p/h_q = t_{pq}, \qquad \text{in } U_p \cap U_q.$$

Proof. It is sufficient to consider one irreducible component, say $\mathbf{\Gamma}_1$, of $\mathbf{\Gamma}$. $\mathbf{\Gamma}_1$ is a compact Riemann surface; hence it can be decomposed into the sum of a finite number of 2-simplices so that $\mathbf{\Gamma}_1$ becomes a finite simplicial complex $\mathfrak{K}$. We denote the vertices (0-simplices) of $\mathfrak{K}$ by $p_1, p_2, \cdots, p_j, \cdots$, the oriented 1-simplex with the vertices p_j, p_k (in order) by S_{jk}, and the oriented 2-simplex with the vertices p_j, p_k, p_l by S_{jkl}. The simplical decomposition $\mathfrak{K}$ of $\mathbf{\Gamma}_1$ determines its dual cell decomposition $\mathfrak{K}^*$ of $\mathbf{\Gamma}_1$. The 2-cell of $\mathfrak{K}^*$ which is dual to p_j will be denoted by E_j. We assume that the simplicial decomposition $\mathfrak{K}$ is cosen so that i) each E_j is contained in one of U_p, say U_{p_j} and ii) each E_j is covered by the domain of the local uniformization parameter τ_j with the center p_j. Now, choosing for each E_j a simply connected neighborhood U_j such that i) $E_j \subset U_j \subset U_{p_j}$, that ii) U_j does not intersect with any simplex S_{klm} not containing the vertex p_j and that iii) each intersection $U_j \cap U_k$ is simply connected (if it is not empty), we put

$$t_{jk} = t_{p_j p_k}, \quad \text{in } U_j \cap U_k. \tag{2.4}$$

Then we have

$$t_{jk} t_{kj} = 1, \qquad \text{in } U_j \cap U_k, \tag{2.5}$$

$$t_{jk} t_{kl} t_{lj} = 1, \qquad \text{in } U_j \cap U_k \cap U_l. \tag{2.6}$$

In order to prove our theorem, it is sufficient to show the existence of a system $\{h_j\}$ of meromorphic functions h_j defined respectively in U_j such that $h_j/h_k = t_{jk}$, in $U_j \cap U_k$; indeed, the desired system $\{h_p\}$ can be obtained from $\{h_j\}$ by putting $h_p = t_{pp_j} h_j$, in $U_p \cap U_j$.

Now, let us forget the definition (2.4) of t_{jk} and consider an *arbitrary* system $\{t_{jk}\}$ of holomorphic functions t_{jk} defined respectively in some neighborhood of $[U_j \cap U_k]$ on $\mathbf{\Gamma}_1$ satisfying the conditions (2.5) and (2.6). We shall say, for a moment, that two such systems $\{t_{jk}\}$ and $\{t'_{jk}\}$ are *equivalent* if t_{jk}/t'_{jk} can be represented as $t_{jk}/t'_{jk} = h_j/h_k$, in $U_j \cap U_k$, by means of a system $\{h_j\}$ of meromorphic functions h_j defined respectively in U_j. Then, it is sufficient for our purpose to show that $\{t_{jk}\}$ is *equivalent* to $\{1\}$. Since $U_j \cap U_k$ is simply connected, the logarithmic function

$$\rho_{jk} = -\rho_{kj} = (1/2\pi i) \log t_{jk} + c_{jk}$$

is determined uniquely in $U_j \cap U_k$ up to an arbitrary rational integer c_{jk}. Putting

$$c_{jkl} = \rho_{jk} + \rho_{kl} + \rho_{lj}, \qquad \text{in } U_j \cap U_k \cap U_l,$$

we get, by virtue of (2.6), rational integers c_{jkl}. These integers c_{jkl} depend on the choice of c_{jk}; but the sum

$$J = \sum_{j<k<l} (\operatorname{sgn} S_{jkl}) \, c_{jkl}$$

of c_{jkl} over all simplices S_{jkl} is independent of the choice of c_{jk}, where sgn S_{jkl} denotes the orientation of the simplex S_{jkl}. This sum J will be called the *characteristic* index of the system $\{t_{jk}\}$ and be denoted by $J\{t_{jk}\}$.

First, we consider the case that $J=0$. In this case, we can choose c_{jk} so that $c_{jkl}=0$. In fact, the relation $J=0$ implies that the 2-cocycle $\sum_{j<k<l} c_{jkl}S_{jkl}$ is the coboundary of a 1-cochain $\sum_{j<k} c'_{jk}S_{jk}$ with integral coefficients c'_{jk}, so that we have $c_{jkl}=c'_{jk}+c'_{kl}+c'_{lj}$. Hence, replacing ρ_{jk} by $\rho'_{jk}=\rho_{jk}-c'_{jk}$, we get $\rho'_{jk}+\rho'_{kl}+\rho'_{lj}=0$. Thus we may assume without loss of generality that

$$\rho_{jk}+\rho_{kl}+\rho_{lj}=0, \qquad \text{in } U_j\cap U_k\cap U_l \tag{2.7}$$

Now, assuming (2.7), we shall prove that there exists a system $\{\eta_j\}$ of *holomorphic* functions η_j defined respectively in U_j such that $\kappa_{jk}=\rho_{jk}-\eta_j+\eta_k$ is constant in $U_j\cap U_k$. To do this, we first remark that $\mathbf{\Gamma}_1$ can be considered as a Kählerian manifold. For an arbitrary function $\psi(\tau,\bar\tau)$ of class C^∞ on $\mathbf{\Gamma}_1$, τ being the local uniformization variables on $\mathbf{\Gamma}_1$, we denote $(\partial\psi/\partial\tau)d\tau$ by $d_\tau\psi$ and $(\partial\psi/\partial\bar\tau)d\bar\tau$ by $\bar d_\tau\psi$. Now, by virtue of (2.7), one can readily construct functions $\psi_j(\tau,\bar\tau)$ of class C^∞ defined respectively in U_j so that[26]

$$\psi_j(\tau,\bar\tau)-\psi_k(\tau,\bar\tau)=\rho_{jk}(\tau), \qquad \text{in } U_j\cap U_k. \tag{2.8}$$

Since $\rho_{jk}(\tau)$ is holomorphic, we have $\bar d_\tau\psi_j=\bar d_\tau\psi_k$ in $U_j\cap U_k$. This shows that $Q=\bar d_\tau\psi_j$ is a well determined 1-form of class C^∞ on $\mathbf{\Gamma}_1$; moreover, since $dQ=(\partial^2\psi_j/\partial\tau\partial\bar\tau)d\tau d\bar\tau$, dQ is a pure 2-form of type 1. Putting $\Theta=(id\Lambda G-\delta G)dQ$, we get therefore a pure 1-form Θ of class C^∞ of type 0 satisfying $d\Theta=-dQ$, as Theorem 1.4 shows. It follows from $d(Q+\Theta)=0$ that there exists in each U_j a function $\alpha_j(\tau,\bar\tau)$ of class C^∞ satisfying

$$d\alpha_j=Q+\Theta. \tag{2.9}$$

Since Q or Θ is a pure 1-form of type 1 or 0, respectively, this yields $\bar d_\tau\alpha_j=Q$, proving that $\bar d_\tau(\psi_j-\alpha_j)=0$. Hence $\eta_j=\psi_j-\alpha_j$ is a holomorphic function in U_j. On the other hand, (2.9) shows that $d(\alpha_j-\alpha_k)=0$. Hence $\kappa_{jk}=\alpha_j-\alpha_k$ is a constant. Inserting $\psi_j=\eta_j+\alpha_j$ in (2.8), we get therefore

$$\rho_{jk}(\tau)=\eta_j(\tau)-\eta_k(\tau)+\kappa_{jk}.$$

Putting $e_j=\exp 2\pi i\eta_j$, $t^0_{jk}=\exp 2\pi i\kappa_{jk}$, we have therefore $t_{jk}=e_j/e_k\cdot t^0_{jk}$. Thus the system $\{t_{jk}\}$ is equivalent to the system $\{t^0_{jk}\}$ consisting of the *constants* t^0_{jk}.

Since $\kappa_{jk}+\kappa_{kl}+\kappa_{lj}=0$, the sum $\sum_{j<k}\kappa_{jk}S_{jk}$ is a 1-cocycle on $\mathfrak{R}$. This

---

[26] Using the fact that each ρ_{jk} is extended analytically on a neighborhood of the closure of $U_j\cap U_k$, we can construct successively $\psi_1,\psi_2,\cdots,\psi_k,\cdots$ so that $\psi_k=\psi_j+\rho_{kj}$ in $U_k\cap U_j$ for $1\leqq j\leqq k-1$. Cf. Weil [24], p. 113.

cocycle determines a *character* χ of the 1-Betti group of $\mathfrak{K}$ over the ring of integers defined as follows:

$$\chi(\gamma) = \exp 2\pi i (\sum_{j<k} \gamma_{jk} \kappa_{jk})$$

for an arbitrary integral 1-cycle $\gamma = \sum_{j<k} \gamma_{jk} S_{jk}$. Now, let f be a multiplicative meromorphic function on $\mathbf{\Gamma}_1$ belonging to the character $\chi(\gamma)$. Then, denoting by f_j an arbitrary branch of f in U_j and putting

$$f_j/f_k = \exp 2\pi i \lambda_{jk}, \qquad (\lambda_{kj} = -\lambda_{jk}),$$

we get $\chi(\gamma) = \exp 2\pi i (\sum_{j<k} \gamma_{jk} \lambda_{jk})$. Hence $\sum_{j<k} \gamma_{jk}(\kappa_{jk} - \lambda_{jk}) \equiv 0 \bmod 1)$ for arbitrary 1-cycles $\gamma = \sum_{j<k} \gamma_{jk} S_{jk}$. This shows that, considered as a 1-cocycle with complex coefficients modulus 1, $\sum_{j<k} (\kappa_{jk} - \lambda_{jk}) S_{jk}$ is cohomologous to zero. Consequently there exists a system $\{\kappa_j\}$ of constants κ_j such that $\kappa_{jk} \equiv \lambda_{jk} + \kappa_j - \kappa_k \pmod 1$. Putting $h_j = f_j \cdot \exp 2\pi i \kappa_j$, we get therefore $t^0_{jk} = h_j/h_k$, in $U_j \cap U_k$, proving that the system $\{t^0_{jk}\}$ is equivalent to $\{1\}$. Hence the system $\{t_{jk}\}$ is also equivalent to $\{1\}$.

Now, we consider the general case that $J \neq 0$. Let $\{t'_{jk}\}$ be the system defined as $t'_{jk}(\tau) = t_{jk}(\tau) \tau_j^{m_j} / \tau_k^{m_k}$, where τ_j is the local uniformization variable with the center p_j and m_j is arbitrary rational integers. Then we get

$$J\{t'_{jk}\} = J\{t_{jk}\} + \sum_j m_j. \tag{2.10}$$

In fact, by the transformation $t_{jk} \to t'_{jk}$, ρ_{jk} is transformed to

$$\rho_{jk}(\tau) = \rho_{jk}(\tau) + (1/2\pi i)\{m_j \log_k \tau_j - m_k \log_j \tau_k\},$$

where $\log_k \tau_j$ denotes a branch of $\log \tau_j$ in $U_j \cap U_k$. Hence we get

$$J\{t'_{jk}\} = J\{t_{jk}\} + (1/2\pi i) \sum_j m_j \sum_{k<l} \operatorname{sgn} S_{jkl} \cdot \{\log_k \tau_j - \log_l \tau_j\},$$

yielding immediately (2.10). By a suitable choice of the integers m_j, we get therefore a system $\{t'_{jk}\}$ with $J\{t'_{jk}\} = 0$ which is equivalent to $\{t_{jk}\}$. Thus our problem is reduced to the special case that $J = 0$, q. e. d.

The system $\{h_p\}$ introduced in the above Theorem 2.1 is not determined uniquely. Firstly, it depends on the choice of the system of minimal local equations $\{R_{\mathfrak{p}} = 0\}$. Let $\{R'_{\mathfrak{p}} = 0\}$ be another system of minimal local equations. Then

$$M_{\mathfrak{p}}(z^1, z^2) = R'_{\mathfrak{p}}(z^1, z^2) / R_{\mathfrak{p}}(z^1, z^2)$$

is a non-vanishing holomorphic function in a neighborhood $\mathfrak{U}_{\mathfrak{p}}$ of $\mathfrak{p}$, and the meromorphic functions h'_p defined by

$$h'_p(\tau) = M_{\phi(p)}(\phi^1(\tau), \phi^2(\tau)) \cdot h_p(\tau) \tag{2.11}$$

constitute a system $\{h'_p\}$ satisfying (2.3) with respect to $\{R'_\mathfrak{p}\}$. Thus, *the system* $\{h_p\}$ *can be considered as a "quantity" which is represented by the "coordinates"* h_p *with respect to the given "reference frame"* $\{R_\mathfrak{p}\}$ *and the "coordinates"* h'_p *corresponding to any other reference frame* $\{R'_\mathfrak{p}\}$ *is linked to* h_p *by the transformation* (2.11) *associated with the transition* $\{R_\mathfrak{p}\}\to\{R'_\mathfrak{p}\}$.[27] Secondly, multiplying each h_p by an arbitrary fixed meromorphic function f on $\mathbf{\Gamma}$, we get another system $\{h'_p \mid h'_p = fh_p\}$ satisfying (2.3) with respect to the same frame $\{R_\mathfrak{p}\}$. Conversely, for arbitrary two systems $\{h_p\}$ and $\{h'_p\}$ satisfying (2.3) with respect the same frame $\{R_\mathfrak{p}\}$, we have $h'_p/h_p = h'_q/h_q$, in $U_p \cap U_q$, proving that $f = h'_p/h_p$ is a well determined meromorphic function on $\mathbf{\Gamma}$. Thus, *for fixed* $\{R_\mathfrak{p}\}$, *the system* $\{h_p\}$ *satisfying* (2.3) *is unique up to the multiplication by an arbitrary meromorphic function* f *on* $\mathbf{\Gamma}$.

For an arbitrary meromorphic function

$$f(\tau_p) = \tau_p{}^m(a_0 + a_1\tau_p + a_2\tau_p{}^2 + \cdots), \qquad (a_0 \neq 0)$$

defined in U_p, we denote by $V_p(f)$ the exponent m appearing in the above expansion of f: $V_p(f) = m$. The functional V_p thus defined is called the *valuation with the center* p. In case f is a meromorphic function defined on the whole of $\mathbf{\Gamma}$, the divisor (f) of f can be written as

$$(f) = \sum_p V_p(f) \cdot p.$$

Corresponding to this, *we define the divisor* $\mathfrak{d}$ *of an arbitrary system* $\{h_\mathfrak{p}\}$ *satisfying* (2.3) *as*

$$\mathfrak{d} = \sum_p V_p(h_p) \cdot p.$$

As (2.11) shows, $\mathfrak{d}$ is independent of the choice of the reference frame $\{R_\mathfrak{p}\}$. *By a characteristic divisor of* Γ *we shall mean the divisor* $\mathfrak{d}$ *of the system* $\{h_p\}$ *introduced in Theorem* 2.1. Let $\mathfrak{d}'$ be the divisor of $\{h'_p\} = \{fh_p\}$; then we have $\mathfrak{d}' - \mathfrak{d} = (f)$. *Thus, the divisor class* $\{\mathfrak{d}\}$ *of* $\mathfrak{d}$ *on* $\mathbf{\Gamma}$ *is determined uniquely by* Γ *and does not depend on the choice of the systems* $\{h_p\}$ *and* $\{R_\mathfrak{p}\}$. *The divisor class* $\{\mathfrak{d}\}$ *will be called the characteristic divisor class*[28] *of* Γ.

The condition (2.3) can be written as

$$h_p(\tau)/h_q(\tau) = (R_p(z)/R_q(z))_{z=\phi(\tau)}. \tag{2.12}$$

---

[27] See Weyl [29], pp. 13-23.

[28] The characteristic divisor class corresponds to the *characteristic series* of Italian geometers. See Remark 2 at the end of this Section.

Denoting $\partial R_p/\partial z^a$ by $\partial_a R_p$, we have therefore

$$(2.13) \qquad \partial_a R_p(\phi(\tau))/h_p(\tau) = \partial_a R_q(\phi(\tau))/h_q(\tau), \qquad \text{in } U_p \cap U_q.$$

THEOREM 2.2. *The degree of the characteristic divisor* $\mathfrak{d}$ *of* Γ *is equal to the topological intersection number* $I(\Gamma, \Gamma)$ *of the cycle* Γ *with itself*: $\deg \mathfrak{d} = I(\Gamma, \Gamma)$.

Proof. Let $\mathfrak{U}_\mathfrak{p}$ be a sufficiently small *spherical* neighborhood of a point $\mathfrak{p} \varepsilon \Gamma$ such that the local equation $R_\mathfrak{p}(z^1, z^2) = 0$ is useful in the whole of $\mathfrak{U}_\mathfrak{p}$; again, let $T \subset \mathfrak{U}_\mathfrak{p}$ be a 2-chain of class C^2 whose boundary $\mathfrak{B}T$ does not meet with Γ. Then, as one readily infers, the intersection number $I(T, \Gamma)$ is given by

$$(2.14) \qquad I(T, \Gamma) = (1/2\pi i) \int_{\mathfrak{B}T} d \log R_\mathfrak{p}(z^1, z^2).$$

Now, let

$$\mathfrak{d} = \sum_{k=1}^{\lambda} m_k p_k - \sum_{k=\lambda+1}^{\lambda+\mu} m_k p_k = \sum_p V_p(h_p) p$$

be the characteristic divisor determined by a system $\{h_p\}$ satisfying (2.3), where m_k are positive integers. We choose $\{h_p\}$ so that no $\mathfrak{p}_k = \phi(p_k)$ $(k = 1, 2, \cdots, \lambda, \cdots, \lambda + \mu)$ is a singular point of Γ. Again, we denote by $\{\mathfrak{p}_{\lambda+\mu+1}, \mathfrak{p}_{\lambda+\mu+2}, \cdots, \mathfrak{p}_{\lambda+\mu+\nu}\}$ the set of all singular points of Γ. We choose for each $\mathfrak{p}_k$ $(k = 1, 2, \cdots, \lambda + \mu + \nu)$ two sufficiently small neighborhoods $\mathfrak{U}_k$, $\mathfrak{U}'_k$ such that $\mathfrak{p}_k \varepsilon \mathfrak{U}'_k \subset \mathfrak{U}_k$ and that $\mathfrak{U}'_k$ is sufficiently small compared with $\mathfrak{U}_k$. Again, for each $\mathfrak{p}_k$ $(k = 1, 2, \cdots, \lambda)$, we choose $\epsilon_k > 0$ so that the neighborhood

$$U_k = \{\tau \mid |h_{p_k}(\tau)| < \epsilon_k\}$$

of p_k satisfies $U_k \supseteq [\phi^{-1}(\mathfrak{U}_k)]$ (remember that $h_{p_k}(\tau)$ vanishes at p_k). Now, consider the covariant vector field $v(\mathfrak{q}) = (v^1(\mathfrak{q}), v^2(\mathfrak{q}))$, $\mathfrak{q} \varepsilon \Gamma$, defined as follows:

$$(2.15) \qquad \begin{cases} v^a(\mathfrak{q}) = \sum_\beta g^{a\beta^*}(\mathfrak{q}) \overline{\partial_\beta R_{p_k}(\mathfrak{q})} \cdot h_{p_k}(q)/\epsilon_k{}^2, \\ \qquad\qquad\qquad \text{for } \mathfrak{q} \varepsilon \phi(U_k) \quad (k = 1, 2, \cdots, \lambda), \\ v^a(\mathfrak{q}) = 0, \qquad \text{for } \mathfrak{q} = \mathfrak{p}_k \quad (k = \lambda + \mu + 1, \cdots, \lambda + \mu + \nu), \\ v^a(\mathfrak{q}) = \sum_\beta g^{a\beta^*}(\mathfrak{q}) \cdot \overline{\partial_\beta R_p(\mathfrak{q})}/\overline{h_p(q)}, \qquad \text{otherwise,} \end{cases}$$

where $q = \phi^{-1}(\mathfrak{q})$. $v(\mathfrak{q})$ is a *continuous* vector field, as one readily infers by means of (2.13) (remember that U_k is defined so that $h_{p_k}\bar{h}_{p_k} = \epsilon_k{}^2$ on the boundary of U_k). We choose $\delta > 0$ sufficiently small and construct a continuously differentiable vector field $u(\mathfrak{p})$ defined on the whole of $\mathfrak{M}$ so that $N_{\mathfrak{q}}[u(\mathfrak{q}) - v(\mathfrak{q})] < \delta$, for all $\mathfrak{q} \varepsilon \Gamma$, where $N_{\mathfrak{q}}[u]$ denotes the norm $\sum g_{\alpha\beta^*}(\mathfrak{q}) u^\alpha \bar{u}^\beta$ of the vector u. $v(\mathfrak{q})$ does not vanish except for $\mathfrak{q} = \mathfrak{p}_k$ $(k = 1, 2, \cdots, \lambda + \mu + \nu)$. Hence, for sufficiently small $\delta > 0$, $u(\mathfrak{q})$ does not vanish for $\mathfrak{q} \varepsilon \Gamma - \mathfrak{U}'_1 - \mathfrak{U}'_2 \cdots - \mathfrak{U}'_{\lambda+\mu+\nu}$. Now, consider the differential equation $dz^\alpha(\mathfrak{p})/dt = u^\alpha(\mathfrak{p})$, $(t \geqq 0)$, $z^\alpha(\mathfrak{p})$ being the local coordinates of the *variable* point $\mathfrak{p} \varepsilon \mathfrak{M}$, and, for each $\mathfrak{q} \varepsilon \Gamma$, denote by $\mathfrak{q}_t$ the solution of this equation satisfying the initial condition: $\mathfrak{q}_0 = \mathfrak{q}$. Then, for fixed t, the mapping $\mathfrak{q} \rightarrow \mathfrak{q}_t$ is a continuous mapping and the image Γ_t of the cycle Γ by this mapping $\mathfrak{q} \rightarrow \mathfrak{q}_t$ is homotopic to Γ. Hence we have $I(\Gamma, \Gamma) = I(\Gamma, \Gamma_t)$. For sufficiently small $t > 0$, Γ_t does *not* intersect with Γ outside of $\mathfrak{U}'_1 \cup \mathfrak{U}'_2 \cup \cdots \cup \mathfrak{U}'_{\lambda+\mu+\nu}$. Hence, in order to find the value of $I(\Gamma, \Gamma_t)$, it is sufficient to compute the intersection number $I_k(\Gamma, \Gamma_t)$ of Γ and Γ_t in each $\mathfrak{U}_k$, $k = 1, 2, \cdots, \lambda + \mu + \nu$. For $k \leqq \lambda + \mu$, Γ consists in $\mathfrak{U}_k$ of only one branch $\Gamma^{(1)}$ $(= \Gamma)$, while, for $k > \lambda + \mu$, Γ consists in $\mathfrak{U}_k$ of a finite number of branches $\Gamma^{(1)}, \Gamma^{(2)}, \cdots$ corresponding to points $p_k^{(1)}, p_k^{(2)}, \cdots, \varepsilon\, \phi^{-1}(\mathfrak{p}_k)$. Obviously we have $\Gamma_t = \sum_r \Gamma_t^{(r)}$ in $\mathfrak{U}_k$ and therefore

$$I_k(\Gamma, \Gamma_t) = \sum_r I_k(\Gamma, \Gamma_t^{(r)}).$$

Let $\mathfrak{S}_k$ be the (spherical) boundary of $\mathfrak{U}_k$. Then the intersection $Z_t = \mathfrak{S}_k \cdot \Gamma_t^{(r)}$ of $\mathfrak{S}_k$ with $\Gamma_t^{(r)}$ is a 1-cycle bounding the closure $[\mathfrak{U}_k \cdot \Gamma_t^{(r)}]$ of the intersection $\mathfrak{U}_k \cdot \Gamma_t^{(r)}$. Hence, using (2. 14), we get

$$I_k(\Gamma, \Gamma_t^{(r)}) = (1/2\pi i) \int_{Z_t} d \log R(z),$$

where $R(z) = R_{\mathfrak{p}_k}(z)$. Let $\mathfrak{q}_t \varepsilon Z_t$ be an arbitrary point on Z_t. Then we have

$$z^\alpha(\mathfrak{q}_t) = z^\alpha(\mathfrak{q}) + t \cdot u^\alpha(\mathfrak{q}) + o(t), \qquad (\mathfrak{q} = \mathfrak{q}_0 \varepsilon \Gamma)$$

and therefore

$$R(\mathfrak{q}_t) = t \sum_\alpha \partial_\alpha R(\mathfrak{q}) \cdot u^\alpha(\mathfrak{q}) + o(t) = t \sum_\alpha \partial_\alpha R(\mathfrak{q}) \cdot v^\alpha(\mathfrak{q}) + \kappa \cdot t\delta + o(t),$$

where $R(\mathfrak{q}_t) = R(z(\mathfrak{q}_t))$ and κ is a constant not depending on $\mathfrak{q}$. Inserting (2. 15) for $v^\alpha(\mathfrak{q})$, we get

$$R(\mathfrak{q}_t) = \begin{cases} t \cdot N(\mathfrak{q}) h_p(q)/\epsilon_k^2 + \kappa t\delta + o(t), & (1 \leqq k \leqq \lambda), \\ t \cdot N(\mathfrak{q})/\overline{h_p(q)} + \kappa t\delta + o(t), & (k > \lambda), \end{cases}$$

where $N(\mathfrak{q}) = \sum g^{\alpha\beta^*}(\mathfrak{q}) \partial_\alpha R(\mathfrak{q}) \overline{\partial_\beta R(\mathfrak{q})}$, $q = \phi^{-1}(\mathfrak{q})$ and $p = p_k$ or $p_k^{(r)}$ according as $k \leqq \lambda + \mu$ or $k > \lambda + \mu$. Obviously $N(\mathfrak{q}) \geqq$ const. > 0 uniformly for $\mathfrak{q}_t \varepsilon Z_t$. Choosing $t > 0$ and $\delta >$ sufficiently small, we get therefore

$$(1/2\pi i)\int_{Z_t} d\log R(\mathfrak{q}_t) = (1/2\pi i)\int_{Z_t} d\log h_p(q) = (1/2\pi i)\int_{Z_0} d\log h_p(q)$$

$$= V_p(h_p) = \begin{cases} m_k & (1 \leqq k \leqq \lambda), \\ -m_k & (\lambda < k \leqq \lambda + \mu), \\ 0 & (\lambda + \mu < k \leqq \lambda + \mu + \nu). \end{cases}$$

Consequently we obtain

$$I(\Gamma, \Gamma_t) = \sum_k \sum_r I(\Gamma, \Gamma_t^{(r)}) = \sum_{k=1}^{\lambda} m_k - \sum_{k=\lambda+1}^{\lambda+\mu} m_k,$$

proving that $I(\Gamma, \Gamma) = \deg \mathfrak{d}$, q. e. d.

Remark 1. In terms of the theory of fibre spaces,[29] the relations between the system $\{h_p\}$, the characteristic divisor $\mathfrak{d}$ and the characteristic index J can be expressed as follows: Consider the system $\{U_q \times \mathfrak{E} \mid q \varepsilon \mathbf{\Gamma}\}$ of the direct products $U_q \times \mathfrak{E}$ of U_q and the *projective line* $\mathfrak{E}$ and identify $p \times \zeta_q$ in $U_q \times \mathfrak{E}$ with $p \times \zeta_r$ in $U_r \times \mathfrak{E}$ if $\zeta_q/\zeta_r = t_{qr}(p)$. Then we get from $\{U_q \times \mathfrak{E}\}$ a compact analytic manifold $\mathfrak{F}$ consisting of all points $P = p \times \zeta_q$ which is obviously a *fibre bundle* over the base space $\mathbf{\Gamma}$. We have the natural mapping $\pi : P = p \times \zeta_q \rightarrow p = \pi(P)$ of $\mathfrak{F}$ onto $\mathbf{\Gamma}$ which is called the *projection.* Obviously $\mathfrak{F}$ is determined uniquely by the system $\{R_\mathfrak{p}\}$. Now, let $\mathfrak{F}'$ be the fibre bundle determined by another system $\{R'_\mathfrak{p}\}$. Then the mapping

$$P = p \times \zeta_q \rightarrow P' = p \times \zeta'_q, \qquad \zeta'_q = \zeta_q \cdot [R'_q/R_q]_{z=\phi(p)}$$

is a *biregular mapping of* $\mathfrak{F}$ *onto* $\mathfrak{F}'$ *satisfying* $\pi(P') = \pi(P)$. Thus $\mathfrak{F}$ and $\mathfrak{F}'$ have the same *analytic structure relative to* $\mathbf{\Gamma}$. *The analytic structure of the fibre bundle* $\mathfrak{F}$ *relative to* $\mathbf{\Gamma}$ *is therefore determined uniquely by the imbedding* ϕ *of* $\mathbf{\Gamma}$ *into* $\mathfrak{M}$. A *cross-section*[30] of $\mathfrak{F}$ is, by definition, a curve C on $\mathfrak{F}$ such that π induces on C a biregular transformation of C onto $\mathbf{\Gamma}$. The subsets $C_0 = \{p \times 0 \mid p \varepsilon \mathbf{\Gamma}\}$ and $C_\infty = \{p \times \infty \mid p \varepsilon \mathbf{\Gamma}\}$ are obviously cross-sections of $\mathfrak{F}$. Now it can be easily verfied that[31] *any system* $\{h_q\}$ *satisfying* (2. 3) *corresponds one-to-one to the cross-section* $C = \{p \times h_q(p) \mid p \varepsilon \mathbf{\Gamma}\}$ *which differs from* C_0 *or* C_∞. By means of this cross-section C, the characteristic divisor $\mathfrak{d} = \sum_q V_q(h_q) \cdot q$ *is represented as* $\mathfrak{d} = \pi(C \cdot C_0 - C \cdot C_\infty)$, where $C \cdot C_0$, $C \cdot C_\infty$ denote the intersection products. The divisor class $\{\mathfrak{d}\}$

---

[29] Cf. Weil [25], Cartan [3]. For the general theory of fibre spaces, see Chern [5].
[30] Weil [25].
[31] Weil [25].

of $\mathfrak{d}$ is an *invariant of the analytic structure* of the fibre bundle $\mathfrak{F}$ relative to $\mathbf{\Gamma}$ and, conversely, the analytic structure of $\mathfrak{F}$ relative to $\mathbf{\Gamma}$ is determined uniquely by $\{\mathfrak{d}\}$.[32] From the topological view point, the characteristic index J is the only invariant of the fibre bundle $\mathfrak{F}$. It follows from our construction of the system $\{h_p\}$ that this invariant J is represented as $J = \deg \mathfrak{d}$.

Remark 2. In case there exists on $\mathfrak{M}$ a meromorphic function F which has each component of Γ as a polar curve of multiplicity 1 (this is the case if $\mathfrak{M}$ is an algebraic surface), the system $\{h_p\}$ is obtained simply by putting $h_p(\tau) = R_p F(\phi_p(\tau))$, so that various things concerning $\{h_p\}$ become somewhat simpler. For example, the characteristic divisor can be defined as the intersection product $\mathfrak{d} = D \cdot \Gamma$ of Γ and $D = (F) + \Gamma$, where (F) denotes the divisor of F; again, since $I(\Gamma, \Gamma) = I(D, \Gamma)$, the relation $\deg \mathfrak{d} = I(\Gamma, \Gamma)$ follows immediately from $\mathfrak{d} = D \cdot \Gamma$. The relation $\mathfrak{d} = D \cdot \Gamma$ indicates that $\mathfrak{d}$ is a member of the characteristic series[33] on Γ if $\mathfrak{d}$ is a positive divisor. The system $\{h_p\}$ is in a certain sense only a substitute for such a function F.

3. An existence theorem. By a divisor on $\mathfrak{M}$ we shall mean a 2-cycle $D = \sum m_\nu \Gamma_\nu$ with integral coefficients m_ν consisting of a finite number of irreducible curves Γ_ν. Each Γ_ν associated with $m_\nu \neq 0$ is called a *component* of D. The set of all divisors on $\mathfrak{M}$ constitutes an additive group. Every meromorphic function F on $\mathfrak{M}$ which is not identically zero determines its divisor D in the well known manner (if such a function ever exists). The divisor of F will be denoted by (F). We shall say that the divisor D' is *linearly equivalent* to D and write $D' \approx D$, if there exists on $\mathfrak{M}$ a meromorphic function F such that $D' - D = (F)$. Obviously the *linear equivalence* thus defined is an equivalence relation and therefore the set of all divisors on $\mathfrak{M}$ can be decomposed into mutually disjoint equivalence classes. Each equivalence class will be called a *divisor class*; the divisor class containing D will be denoted by $\{D\}$. We shall say that $D = \sum m_\nu \Gamma_\nu$ is *positive* and write $D > 0$, if all coefficients m_ν are $\geqq 0$ and $D \neq 0$. A meromorphic function F on $\mathfrak{M}$ will be called a *multiple* of D if $(F) - D \geqq 0$ or F is identically zero. The set of all meromorphic functions which are multiples of $-D$ will be denoted by $\mathfrak{F}(D)$:

$$\mathfrak{F}(D) = \{F \mid (F) + D \geqq 0\}.$$

$\mathfrak{F}(D)$ is a linear space over the field of all complex numbers. By the *dimension* of the divisor class $\{D\}$ we shall mean the dimension of the linear space $\mathfrak{F}(D)$. This definition is legitimate, since the dimension of $\mathfrak{F}(D)$ is determined uniquely by the divisor class $\{D\}$ of D. The dimension of $\{D\}$ will be denoted by $\dim\{D\}$, i. e. we put $\dim\{D\}=\dim\mathfrak{F}(D)$. Clearly an arbitrary curve $\Gamma=\sum_{\nu=1}^{\mu}\Gamma_\nu$ mentioned in Section 2 can be considered as a divisor. Our main purpose is to estimate the dimension of the divisor class $\{\Gamma\}$ of an arbitrary curve Γ.

---

[32] Weil [25].

[33] Cf. Zariski [30], p. 26.

Let Γ, $R_{\mathfrak{p}}$, $\phi_p{}^a$, h_p, etc. have the same meaning as in *Section* 2. First we shall show that *the meromorphic function* $1/R_{\mathfrak{p}}$ *can be considered as a* 0-*current defined in* $\mathfrak{U}_{\mathfrak{p}}$, where $\mathfrak{U}_{\mathfrak{p}}$ is a sufficiently small neighborhood of $\mathfrak{p}$. For this purpose, we take an arbitrary 4-form Φ of class C^∞ with $|\Phi|\subset\mathfrak{U}_{\mathfrak{p}}$ and consider the integral $\int(1/R_{\mathfrak{p}})\cdot\Phi$. *In case* $\mathfrak{p}$ *is a simple point of* Γ, *the integral* $\int(1/R_{\mathfrak{p}})\cdot\Phi$ *converges absolutely*; then we put

$$(3.1)_0 \qquad (1/R_{\mathfrak{p}})[\Phi]=\int(1/R_{\mathfrak{p}})\cdot\Phi.$$

In case $\mathfrak{p}$ is a singular point of Γ, we define the "*principal value*" $(1/R_{\mathfrak{p}})[\Phi]$ of the (divergent) integral $\int(1/R_{\mathfrak{p}})\cdot\Phi$ as follows: Choosing the system of local coordinates (w,z) with the *origin* $\mathfrak{p}$ (we write w, z instead of z^1,z^2) so that $R_{\mathfrak{p}}(w,0)$ does not vanish identically, we put

$$(3.1) \qquad (1/R_{\mathfrak{p}})[\Phi]=\lim_{\epsilon\to+0}\int_{|z|>\epsilon}(1/R_{\mathfrak{p}})\cdot\Phi.$$

The *existence of the limit* in the right hand side of (3.1) is proved as follows: $\phi^{-1}(\mathfrak{p})$ consists of a finite number of points $p_1,p_2,\cdots,p_l$ corresponding to the irreducible branches $\Gamma_1,\Gamma_2,\cdots,\Gamma_l$ of Γ passing through $\mathfrak{p}$. Corresponding to this, $R_{\mathfrak{p}}(w,z)$ is decomposed into a product

$$(3.2) \quad R_{\mathfrak{p}}(w,z)=M(w,z)R_1(w,z)R_2(w,z)\cdots R_l(w,z), \qquad (M(0,0)\neq 0)$$

of the "unit" M and the irreducible distinguished polynomials[34] $R_1(w,z),\cdots,R_l(w,z)$ in w representing respectively the irreducible branches $\Gamma_1,\cdots,\Gamma_l$; each $R_s(w,z)$ is written therefore as

$$R_s(w,z)=w^{m_s}+B_{s1}(z)w^{m_s-1}+\cdots+B_{sm_s}(z),$$

where $B_{s\nu}(z)$ are holomorphic functions of z satisfying $B_{s\nu}(0)=0$. For our present purpose we may assume that $M(w,z)=1$, since the integral in the right hand side of (3.1) can be rewritten as

$$\int_{|z|>\epsilon}(1/\prod_s R_s)\cdot\Phi' \qquad\qquad \Phi'=M^{-1}\Phi.$$

---

[34] See Behnke and Thullen [1], p. 57, Bochner and Martin [2], Chap. IX.

Under these assumptions, we can find such a parametric representation of each Γ_s that

$$\begin{cases} w = \mathfrak{P}_s(\tau_s) = \tau_s^{k_s}(a_{s0} + a_{s1}\tau_s + \cdots), & (a_{s0} \neq 0,\ k_s \geqq 1), \\ z = \tau_s^{m_s}, & \end{cases}$$

so that $R_{\mathfrak{p}}(w, z)$ is decomposed as

$$(3.3) \qquad R_{\mathfrak{p}}(w, z) = \prod_s \prod_{k=0}^{m_s-1} [w - \mathfrak{P}_s(\zeta_s^k \tau_s)], \qquad (\zeta_s = \exp(2\pi i/m_s)).$$

This yields immediately

$$(3.4) \qquad 1/R_{\mathfrak{p}}(w, z) = \sum_s \sum_{k=0}^{m_s-1} \rho_s(\zeta_s^k \tau_s)[w - \mathfrak{P}_s(\zeta_s^k \tau_s)]^{-1},$$

where

$$\rho_s(\tau_s) = 1/\partial_1 R_{\mathfrak{p}}(\mathfrak{P}_s(\tau_s), \tau_s^{m_s}), \qquad (\partial_1 R_{\mathfrak{p}} = \partial R_{\mathfrak{p}}/\partial w).$$

Now, putting

$$(3.5) \qquad \psi(v, z) = \int \Phi_{11^*22^*}(v + w, z) w^{-1} dw d\bar{w}$$

and $\epsilon_s = \epsilon^{1/m_s}$, we get, using (3.4),

$$\begin{aligned} &\int_{|z| > \epsilon} (1/R_{\mathfrak{p}}) \cdot \Phi \\ &\quad = \sum_s \sum_k \int_{|\tau_s| > \epsilon_s} \rho_s(\zeta_s^k \tau_s) \psi(\mathfrak{P}_s(\zeta_s^k \tau_s), \tau_s^{m_s}) \, | \tau_s |^{2m_s - 2} m_s d\tau_s d\bar{\tau}_s \\ &\quad = \sum_s \int_{|\tau_s| > \epsilon_s} \rho_s(\tau_s) \psi(\mathfrak{P}_s(\tau_s), \tau_s^{m_s}) \, | \tau_s |^{2m_s - 2} m_s^2 \, d\tau_s d\bar{\tau}_s, \end{aligned}$$

since

$$\int \Phi_{11^*22^*}(w, \tau_s^{m_s}) [w - \mathfrak{P}_s(\tau_s)]^{-1} dw d\bar{w} = \int \Phi_{11^*22^*}(w + \mathfrak{P}_s(\tau_s), \tau_s^{m_s}) w^{-1} dw d\bar{w}.$$

As the integral in (3.5) converges absolutely, $\psi(v, z)$ is a function of class C^∞ of $v, \bar{v}, z, \bar{z}$. Consequently $\psi(\mathfrak{P}_s(\tau_s), \tau_s^{m_s})$ is a function of class C^∞ of τ_s, $\bar{\tau}_s$ and therefore *the limit*

$$\lim_{\epsilon \to +0} \int_{|\tau_s| > \epsilon} \rho_s(\tau_s) \psi(\mathfrak{P}_s(\tau_s), \tau_s^{m_s}) \, | \tau_s |^{2m_s - 2} d\tau_s d\bar{\tau}_s$$

does exist; indeed this limit represents the *Cauchy principal value* of the (possibly divergent) integral

$$\int \rho_s \psi(\mathfrak{P}_s, \tau_s^{m_s}) \, | \tau_s |^{2m_s - 2} d\tau_s d\bar{\tau}_s.$$

Thus we infer the existence of the limit in the right-hand side of (3.1); moreover it follows from the above result that *the limit is given by*

$$(3.6) \qquad (1/R_{\mathfrak{p}})[\Phi] = \sum_{s} \lim_{\epsilon \to +0} \int_{|\tau_s| > \epsilon} \rho_s \psi(\mathfrak{P}_s, \tau_s{}^{m_s}) \mid \tau_s \mid^{2m_s - 2} m_s{}^2 d\tau_s d\bar{\tau}_s.$$

By using (3.5) and (3.6), it can be easily verified that the linear functional $(1/R_{\mathfrak{p}})[\Phi]$ defined by (3.1) satisfies the continuity condition for currents. Indeed, it follows from (3.5) that

$$\mid \partial^n \psi / \partial v^h \partial \bar{v}^i \partial z^k \partial \bar{z}^j \mid \leqq c \max \mid \partial^n \Phi_{11^* 22^*} / \partial w^h \partial \bar{w}^i \partial z^k \partial \bar{z}^j \mid,$$

c being a constant which is independent of Φ. Consequently, if $\{\Phi^{(m)}\}$ is a sequence of 4-forms of class C^∞ with $\mid \Phi^{(m)} \mid \subset \mathfrak{U}_{\mathfrak{p}}$ such that each partial derivative $\partial^n \Phi^{(m)}{}_{11^* 22^*} / \partial w^h \partial \bar{w}^i \partial z^k \partial \bar{z}^j$ converges uniformly to zero for $m \to \infty$, we get, by (3.6), $\lim_{m \to \infty} (1/R_{\mathfrak{p}})[\Phi^{(m)}] = 0$. Hence, identifying $1/R_{\mathfrak{p}}$ with the linear functional $(1/R_{\mathfrak{p}})[\Phi]$, we can consider $1/R_{\mathfrak{p}}$ as a 0-current defined in $\mathfrak{U}_{\mathfrak{p}}$. As will be proved below, *the* 0-*current* $1/R_{\mathfrak{p}}$ *thus defined is independent of the choice of the system of local coordinates* (w, z) *appearing in the definition* (3.1).

Putting $\mathfrak{Q}_{\mathfrak{p}} = (1/4\pi i)(1 + iC) d(1/R_{\mathfrak{p}})$, we introduce the 1-*current* $\mathfrak{Q}_{\mathfrak{p}}$ in $\mathfrak{U}_{\mathfrak{p}}$. Since $d(1/R_{\mathfrak{p}})$ is holomorphic in $\mathfrak{U}_{\mathfrak{p}} - \Gamma$, we have $\mid \mathfrak{Q}_{\mathfrak{p}} \mid \subset \Gamma$. Again $\mathfrak{Q}_{\mathfrak{p}}$ satisfies $C\mathfrak{Q}_{\mathfrak{p}} = - i\mathfrak{Q}_{\mathfrak{p}}$, showing that $\mathfrak{Q}_{\mathfrak{p}}$ *is a pure* 1-*current of type* 1. Let $N_{\mathfrak{p}}$ be an arbitrary holomorphic function in $\mathfrak{U}_{\mathfrak{p}}$. Then the meromorphic function $N_{\mathfrak{p}}/R_{\mathfrak{p}}$ is considered as a 0-current, since it can be interpreted as a product of the 0-current $1/R_{\mathfrak{p}}$ and the 0-form $N_{\mathfrak{p}}$. We have

$$d(N_{\mathfrak{p}}/R_{\mathfrak{p}}) = (1/R_{\mathfrak{p}}) \cdot dN_{\mathfrak{p}} + N_{\mathfrak{p}} \cdot d(1/R_{\mathfrak{p}}),$$

while the holomorphic 1-form $dN_{\mathfrak{p}}$ satisfies $(1 + iC) dN_{\mathfrak{p}} = 0$. Hence we get

$$(3.7) \qquad (1 + iC) d(N_{\mathfrak{p}}/R_{\mathfrak{p}}) = 4\pi i \mathfrak{Q}_{\mathfrak{p}} \cdot N_{\mathfrak{p}}.$$

Similarly, for an arbitrary holomorphic double differential $B_{\mathfrak{p}} = B_{\mathfrak{p}12} dz^1 dz^2$ in $\mathfrak{U}_{\mathfrak{p}}$, the double differential $B_{\mathfrak{p}}/R_{\mathfrak{p}}$ is considered as a 2-current in $\mathfrak{U}_{\mathfrak{p}}$. We have $d(B_{\mathfrak{p}}/R_{\mathfrak{p}}) = d(1/R_{\mathfrak{p}}) \cdot B_{\mathfrak{p}}$, while, since $B_{\mathfrak{p}}/R_{\mathfrak{p}}$ is a pure 2-current of type 0, $d(B_{\mathfrak{p}}/R_{\mathfrak{p}})$ is a pure 3-current of type 1. Hence we get

$$\begin{aligned} 2d(B_{\mathfrak{p}}/R_{\mathfrak{p}}) &= (1 - iC) d(B_{\mathfrak{p}}/R_{\mathfrak{p}}) = (1 - iC)\{d(1/R_{\mathfrak{p}}) \cdot B_{\mathfrak{p}}\} \\ &= \{(1 + iC) d(1/R_{\mathfrak{p}})\} \cdot B_{\mathfrak{p}} = 4\pi i \mathfrak{Q}_{\mathfrak{p}} \cdot B_{\mathfrak{p}}, \end{aligned}$$

proving that

$$(3.8) \qquad d(B_{\mathfrak{p}}/R_{\mathfrak{p}}) = 2\pi i \mathfrak{Q}_{\mathfrak{p}} \cdot B_{\mathfrak{p}}.$$

Our next task is to find an explicit expression of the 1-current $\mathfrak{Q}_\mathfrak{p}$. Since $R_p(\phi_p^1(\tau), \phi_p^2(\tau))$ vanishes identically, we have

$$\partial_1 R_p(\phi_p) \cdot d\phi_p^1 + \partial_2 R_p(\phi_p) \cdot d\phi_p^2 = 0,$$

where $\partial_\alpha = \partial/\partial z^\alpha$ $(\alpha = 1, 2)$ and $\phi_p = (\phi_p^1, \phi_p^2)$. Consequently, putting

$$\sigma_p = \partial\phi_p^2/\partial_1 R_p(\phi_p) = -\partial\phi_p^1/\partial_2 R_p(\phi_p),$$

we can define a *differential* σ_p defined in a neighborhood $U_p \subset \boldsymbol{\Gamma}$ of p. In case $\mathfrak{p} = \phi(p)$ is a simple point of Γ, σ_p is a holomorphic differential.

THEOREM 3.1. *For an arbitrary 3-form Ψ of class C^∞ with $|\Psi| \subset \mathfrak{U}_\mathfrak{p}$, we have*

$$\mathfrak{Q}_\mathfrak{p}[\Psi] = \sum_{\phi(p)=\mathfrak{p}} \lim_{\epsilon\to+0} \int_{|\tau_p|>\epsilon} \sum_\beta \Psi_{12\beta^*}(\phi_p)\sigma_p d\bar{\phi}_p^\beta, \tag{3.9}$$

where $\sum_{\phi(p)=\mathfrak{p}}$ denotes the sum extended over all points $p \varepsilon \phi^{-1}(\mathfrak{p})$.

The proof of this theorem will be given below. Using (3.8), we get from (3.9) immediately the following

COROLLARY. *Let $B_\mathfrak{p} = B_{12}(z)dz^1dz^2$ be a holomorphic double differential in $\mathfrak{U}_\mathfrak{p}$. Then, for an arbitrary 1-form X of class C^∞ with $|X| \subset \mathfrak{U}_\mathfrak{p}$, we have*

$$d(B_\mathfrak{p}/R_\mathfrak{p})[X] = 2\pi i \sum_{\phi(p)=\mathfrak{p}} \lim_{\epsilon\to+0} \int_{|\tau_p|>\epsilon} B_{12}(\phi_p)\sigma_p \cdot X, \tag{3.10}$$

where the symbol X in the integral sign stands for the differential

$$\sum_\beta \{X_\beta(\phi_p)d\phi_p^\beta + X_{\beta^*}(\phi_p)d\bar{\phi}_p^\beta\}$$

on Γ_p induced by X.

It is to be noted here that the expressions (3.9), (3.10) are independent of the choice of local uniformization parameters τ_p, since the limits of the integrals in the right-hand sides of (3.9), (3.10) represent the Cauchy principal values of the integrals. For an arbitrary meromorphic differential ξ, defined in U_p, we denote by $\mathrm{Res}_p[\xi]$ the residue $(1/2\pi i)\oint_p \xi$ of ξ at p, where $\oint_p$ denotes the integral extended over a sufficiently small circle $|\tau_p| = \epsilon$ with the center p. Then we have

THEOREM 3.2. *For an arbitrary holomorphic double differential $B_\mathfrak{p} = B_{12}(z)dz^1dz^2$, we have*

$$\sum_{\phi(p)=\mathfrak{p}} \mathrm{Res}_p[B_{12}(\phi_p)\sigma_p] = 0. \tag{3.11}$$

In order to prove the above two Theorems, we first verify the formula (3. 9) for an arbitrary *simple* point $\mathfrak{p}$ of Γ. Put, for simplicity's sake, $R = R_{\mathfrak{p}}$. Then at least one of the first partial derivatives of R, say $\partial_1 R$, does not vanish at $\mathfrak{p}$. Using (1. 3) and (1. 6), we get

$$\begin{aligned} 4\pi i \mathfrak{Q}_{\mathfrak{p}}[\Psi] &= (1+iC)\,d(1/R)\,[\Psi] \\ &= -(1/R)\,[d(1-iC)\Psi] = -\int (1/R)\,d(1-iC)\Psi, \end{aligned}$$

while, since $d(1/R)\cdot(1-iC)\Psi$ vanishes identically in $\mathfrak{U}_{\mathfrak{p}}-\Gamma$, we have

$$(1/R)\cdot d(1-iC)\Psi = d\{(1/R)(1-iC)\Psi\}, \qquad \text{in } \mathfrak{U}_{\mathfrak{p}}-\Gamma.$$

Consequently we get

$$\begin{aligned} 4\pi i \mathfrak{Q}_{\mathfrak{p}}[\Psi] &= -\lim_{\epsilon\to+0}\int_{|R|=\epsilon} d\{(1/R)(1-iC)\Psi\} = \lim_{\epsilon\to+0}\int_{|R|=\epsilon}(1/R)(1-iC)\Psi \\ &= \lim_{\epsilon\to+0}\int_{|R|=\epsilon} 2(dR/R)\{\Psi_{122^*} - \overline{(\partial_2 R}/\overline{\partial_1 R)}\Psi_{121^*}\}(1/\partial_1 R)\,dz^2 d\bar{z}^2 \\ &= 4\pi i\int \{\Psi_{122^*} - \overline{(\partial_2 R}/\overline{\partial_1 R)}\Psi_{121^*}\}(1/\partial_1 R)\,dz^2 d\bar{z}^2. \end{aligned}$$

Inserting $z^2 = \phi_p{}^2(\tau_p)$ in this expression, we obtain

$$\mathfrak{Q}_{\mathfrak{p}}[\Psi] = \int_\Gamma \sum_\beta \Psi_{12\beta^*}(\phi_p)\sigma_p d\bar{\phi}_p{}^\beta;$$

thus (3. 9) is proved for the simple point $\mathfrak{p}$ of Γ, so that (3. 10) is also proved for the simple point $\mathfrak{p}$.

To prove the formula (3. 11), it is sufficient to consider the case that $\mathfrak{p}$ is a singular point of Γ. Let $\chi = \chi(z)$ be a function of class C^∞ with $|\chi| \subset \mathfrak{U}_{\mathfrak{p}}$ such that $\chi(z) = 1$ in a sufficiently small neighborhood $\mathfrak{V}_{\mathfrak{p}} \subset \mathfrak{U}_{\mathfrak{p}}$ of $\mathfrak{p}$. Then we have $|d\chi| \subset \mathfrak{U}_{\mathfrak{p}} - \mathfrak{p}$, while Γ has no singular point in $\mathfrak{U}_{\mathfrak{p}} - \mathfrak{p}$. Hence, using the formula (3. 10) *proved already for any simple point*, we get

$$d(B_{\mathfrak{p}}/R_{\mathfrak{p}})\,[d\chi] = \sum_{\phi(p)=\mathfrak{p}} 2\pi i\int_{U_p} B_{12}(\phi_p)\sigma_p d\chi(\phi_p).$$

Now, $B_{12}(\phi_p)\sigma_p$ is a holomorphic differential on $U_p - p$ and $\chi(\phi_p) = 1$ in the small neighborhood $U_p \cap \phi_p{}^{-1}(\mathfrak{V}_{\mathfrak{p}})$ of p. Hence we get

$$\begin{aligned} \int_{U_p} B_{12}(\phi_p)\sigma_p d\chi(\phi_p) &= -\int_{U_p} d\{B_{12}(\phi_p)\sigma_p\chi(\phi_p)\} \\ &= \oint_p B_{12}(\phi_p)\sigma_p = 2\pi i\,\mathrm{Res}_p[B_{12}(\phi_p)\sigma_p] \end{aligned}$$

and consequently

$$2\pi i \sum_{\phi(p)=\mathfrak{p}} \operatorname{Res}_p[B_{12}(\phi_p)\sigma_p] = -(B_{\mathfrak{p}}/R_{\mathfrak{p}})[dd\chi] = 0,$$

proving (3.11).

Now, using (3.11), we shall prove (3.9) for an arbitrary singular point $\mathfrak{p}$ of Γ. For this purpose, we first prove the following

LEMMA 3.1. *Let* $\{\xi_p \mid p \in \phi^{-1}(\mathfrak{p})\}$ *be a system of the differentials* ξ_p *defined respectively on the irreducible branches* Γ_p, $p \in \phi^{-1}(\mathfrak{p})$, *such that*

$$\sum_{\phi(p)=\mathfrak{p}} \operatorname{Res}_p[B(\phi_p^1, \phi_p^2)\xi_p] = 0$$

for every holomorphic function $B(z^1, z^2)$ *defined in* $\mathfrak{U}_{\mathfrak{p}}$, $\mathfrak{p}$ *being a singular point of* Γ. *Then, for an arbitrary function* $\psi(z^1, z^2)$ *of class* C^∞ *defined in* $\mathfrak{U}_{\mathfrak{p}}$, *we have*

$$\sum_{\phi(p)=\mathfrak{p}} \lim_{\epsilon\to+0} \int_{|\tau_p|=\epsilon} \psi(\phi_p^1, \phi_p^2)\xi_p = 0.$$

Proof. Let n be a positive integer such that $\tau_p^n \xi_p$ is holomorphic for each $p \in \phi^{-1}(\mathfrak{p})$. The function $\psi(z^1, z^2)$ can be represented in $\mathfrak{U}_{\mathfrak{p}}$ as

$$\psi(z^1, z^2) = P(z^1, z^2, \bar{z}^1, \bar{z}^2) + O(|z|^n),$$

where $P(z^1, z^2, \bar{z}^1, \bar{z}^2)$ is a polynomial in $z^1, z^2, \bar{z}^1, \bar{z}^2$ of degree $\leqq n-1$ and $|z| = (|z^1|^2 + |z^2|^2)^{\frac{1}{2}}$. Hence we have

$$\lim_{\epsilon\to+0} \int_{|\tau_p|=\epsilon} \psi \cdot \xi_p = \lim_{\epsilon\to+0} \int_{|\tau_p|=\epsilon} P(\phi_p^1, \phi_p^2, \bar{\phi}_p^1, \bar{\phi}_p^2) \cdot \xi_p$$

while it is clear that

$$\lim_{\epsilon\to+0} \int_{|\tau_p|=\epsilon} \tau^k \bar{\tau}^m d\tau = 0, \quad \text{for } m = 1, 2, \cdots, \quad k = 0, \pm 1, \pm 2, \cdots.$$

Putting $P(z^1, z^2) = P(z^1, z^2, 0, 0)$, we get therefore

$$\begin{aligned}\sum_{\phi(p)=\mathfrak{p}} \lim_{\epsilon\to+0} \int_{|\tau_p|=\epsilon} \psi \cdot \xi_p &= \sum_{\phi(p)=\mathfrak{p}} \lim_{\epsilon\to+0} \int_{|\tau_p|=\epsilon} P(\phi_p^1, \phi_p^2) \cdot \xi_p \\ &= \sum_{\phi(p)=\mathfrak{p}} 2\pi i \operatorname{Res}_p[P(\phi_p^1, \phi_p^2) \cdot \xi_p] = 0,\end{aligned}$$

q. e. d.

Now, supposing that $R_{\mathfrak{p}}(w, z)$ is represented as (3.3), we put

$\phi_s^1 = \phi_{p_s}^1 = \mathfrak{P}_s(\tau_s)$, $\phi_s^2 = \phi_{p_s}^2 = \tau_s^{m_s}$, $\sigma_s = \sigma_{p_s} = \rho_s(\tau_s) m_s \tau_s^{m_s-1} d\tau_s$, where $\rho_s(\tau_s) = 1/\partial_1 R_{\mathfrak{p}}(\phi_s)$. Then, putting $R = R_{\mathfrak{p}}$, we get

$$4\pi i \mathfrak{Q}_{\mathfrak{p}}[\Psi] = -(1/R)[d(1-iC)\Psi] = -\lim_{\epsilon\to+0}\int_{|z|>\epsilon} d\{(1/R)(1-iC)\Psi\}$$

$$= -\lim_{\epsilon\to+0}\lim_{\delta\to+0}\int_{|z|>\epsilon,\,|R|>\delta} d\{(1/R)(1-iC)\Psi\}$$

$$= -\lim_{\epsilon\to+0}\lim_{\delta\to+0}\Big\{\int_{|z|=\epsilon,\,|R|>\delta}(1/R)(1-iC)\Psi + \int_{|z|>\epsilon,\,|R|=\delta}(1/R)(1-iC)\Psi\Big\},$$

while, as was already proved, we have

$$\lim_{\delta\to+0}\int_{|z|>\epsilon,\,|R|=\delta}(1/R)(1-iC)\Psi = 4\pi i\sum_s\int_{|\tau_s|>\epsilon_s}\sum_\beta \Psi_{12\beta^*}(\phi_s)\sigma_s d\bar{\phi}_s{}^\beta,$$

where $\epsilon_s = \epsilon^{1/m_s}$. Hence putting $\eta(\epsilon) = \int_{|z|=\epsilon}(1/R)(1-iC)\Psi$,

$$\mathfrak{Q}_{\mathfrak{p}}[\Psi] = (1/4\pi i)\lim_{\epsilon\to+0}\eta(\epsilon) + \sum_s\lim_{\epsilon\to+0}\int_{|\tau_s|>\epsilon}\sum_\beta\Psi_{12\beta^*}(\phi_s)\sigma_s d\bar{\phi}_s{}^\beta.$$

Consequently, it is sufficient for our purposes to show that $\lim_{\epsilon\to+0}\eta(\epsilon) = 0$. Since $dzd\bar{z} = 0$ on the cylindrical hypersurface $|z| = \epsilon$, we have

$$\eta(\epsilon) = \int_{|z|=\epsilon}(1/R)2\Psi_{211^*}(w,z)\,dzdwd\bar{w}.$$

Putting $\psi(v,z) = \int 2\Psi_{211^*}(w+v,z)w^{-1}dwd\bar{w}$, we get, in the same way as in the proof of (3. 6), $\eta(\epsilon) = \sum_s\sum_k\int_{|\tau_s|=\epsilon_s}\psi(\mathfrak{P}_s(\zeta_s{}^k\tau_s),\tau_s{}^{m_s})\rho_s(\zeta_s{}^k\tau_s)\tau_s{}^{m_s-1}d\tau_s$ $= \sum_s\int_{|\tau_s|=\epsilon_s}\psi(P_s(\tau_s),\tau_s{}^{m_s})\sigma_s$, showing that

$$\lim_{\epsilon\to+0}\eta(\epsilon) = \sum_s\lim_{\epsilon\to+0}\int_{|\tau_s|=\epsilon}\psi(\phi_s{}^1,\phi_s{}^2)\sigma_s.$$

$\psi(v,z)$ is obviously a function of class C^∞. Hence, by virtue of Lemma 3. 1, we infer from (3. 11) that $\lim_{\epsilon\to+0}\eta(\epsilon) = 0$. Thus the formula (3. 9) is completely proved.

Now, we shall show that *the* 0-*current* $1/R_{\mathfrak{p}}$ *is determined uniquely* in the sense that it is indepedent of the choice of the local coordinates w, z appearing in the definition (3. 1). This is obvious if $\mathfrak{p}$ is a simple point of Γ. In case $\mathfrak{p}$ is a singular point, $1/R_{\mathfrak{p}}$ is therefore determined uniquely in $\mathfrak{U}_{\mathfrak{p}} - \mathfrak{p}$, showing that $\mathfrak{Q}_{\mathfrak{p}}$ is also determined uniquely in $\mathfrak{U}_{\mathfrak{p}} - \mathfrak{p}$. Let $\chi = \chi(z)$ be an arbitrary function of class C^∞ with $|\chi| \subset \mathfrak{U}_{\mathfrak{p}} - \mathfrak{p}$, then we have, by (3. 9),

$$\mathfrak{Q}_{\mathfrak{p}}[\chi\Psi] = \sum_{\phi(p)=\mathfrak{p}}\int_{U_p}\chi(\phi_p)\sum_\beta\Psi_{12\beta^*}(\phi_p)\sigma_p d\bar{\phi}_p{}^\beta.$$

Combined with the uniqueness of $\mathfrak{Q}_{\mathfrak{p}}$ in $\mathfrak{U}_{\mathfrak{p}} - \mathfrak{p}$, this shows that, in $\mathfrak{U}_{\mathfrak{p}} - \mathfrak{p}$, the 2-form $\sum_{\beta} \Psi_{12\beta^*}(\phi_p)\sigma_p d\bar{\phi}_p{}^{\beta}$ is independent of the choice of the local coordinates w, z, while (3.9) proves that $\mathfrak{Q}_{\mathfrak{p}}[\Psi]$ is determined uniquely by the behaviors of $\sum_{\beta} \Psi_{12\beta^*}(\phi_p)\sigma_p d\bar{\phi}_p{}^{\beta}$ in $U_p - p$, $p \,\varepsilon\, \phi^{-1}(\mathfrak{p})$. Hence we infer that the 1-current $\mathfrak{Q}_{\mathfrak{p}}$ is determined uniquely and is independent of the choice of local coordinates w, z. The 0-current $1/R_{\mathfrak{p}}$ is connected with $\mathfrak{Q}_{\mathfrak{p}}$ by the relation $(1 + iC)d(1/R_{\mathfrak{p}}) = 4\pi i\mathfrak{Q}_{\mathfrak{p}}$. In order to deduce the uniqueness of $1/R_{\mathfrak{p}}$ from the uniqueness of $\mathfrak{Q}_{\mathfrak{p}}$, it is sufficient therefore to show that *there exists in* $\mathfrak{U}_{\mathfrak{p}}$ *only one 0-current* T *satisfying the conditions*

$$(1 + iC)dT = 4\pi i\mathfrak{Q}_{\mathfrak{p}}, \qquad T - 1/R_{\mathfrak{p}} = 0, \quad \text{in } \mathfrak{U}_{\mathfrak{p}} - \boldsymbol{\Gamma}.$$

Let T, T' be any two 0-currents satisfying the above two conditions. Then the difference $S = T - T'$ vanishes in $\mathfrak{U}_{\mathfrak{p}} - \boldsymbol{\Gamma}$ and satisfies $CdS = idS$; thus dS is a pure 1-current of type 0, while it is obvious that $ddS = 0$. Hence, by Theorem 1.3, dS must be a holomorphic differential in $\mathfrak{U}_{\mathfrak{p}}$. Consequently dS vanishes in the whole of $\mathfrak{U}_{\mathfrak{p}}$ and therefore S must be a constant. Hence S must vanish in the whole of $\mathfrak{U}_{\mathfrak{p}}$, proving that $T = T'$, q. e. d.

We introduce here the notion of the *residues* of double differentials due to H. Poincaré.[35] Let $W = W_{12}(z)dz^1dz^2$ be a double differential defined in a domain $\mathfrak{D} \subseteqq \mathfrak{M}$ such that, for every point $\mathfrak{p} \,\varepsilon\, \mathfrak{D}$, $R_{\mathfrak{p}}W$ is holomorphic at $\mathfrak{p}$. Then, denoting by $R_{\mathfrak{p}}W_{12}(\phi_p(\tau_p))$ the holomorphic function of τ_p obtained from the holomorphic function $R_{\mathfrak{p}}W_{12}$ of z^1, z^2 by setting $z^{\alpha} = \phi_p{}^{\alpha}(\tau_p)$ $(\alpha = 1, 2)$, we infer from (3.10) that, for each $\mathfrak{p} \,\varepsilon\, \mathfrak{D}$,

$$dW[X] = 2\pi i \sum_{\phi(p)=\mathfrak{p}} \int R_pW_{12}(\phi_p(\tau_p))\sigma_p X, \tag{3.12}$$

X being an arbitrary 2-form of class C^{∞} with $| X | \subset \mathfrak{U}_{\mathfrak{p}}$, where the value of the integral in the righ-hand side is to be interpreted as the Cauchy principal value, in case it does not converge absolutely.[36] (3.12) shows that *the differential* $R_pW_{12}(\phi_p)\sigma_p$ *defined in* U_p *is independent of the choice of* $R_{\mathfrak{p}}$ *and the local coordinates* z^1, z^2, U_p being the neighborhood of p introduced in Section 2. Hence we have

$$R_pW_{12}(\phi_p)\sigma_p = R_qW_{12}(\phi_q)\sigma_q, \qquad \text{in } U_p \cap U_q, \tag{3.13}$$

proving that $\xi = R_pW_{12}(\phi_p)\sigma_p$ is a well defined differential on $\boldsymbol{\Gamma} \cap \phi^{-1}(\mathfrak{D})$

---

[35] Poincaré [17].

[36] (3.12) is essentially the same as a formula due to Poincaré; see Poincaré [17], p. 350.

determined uniquely by W. This differential ξ is called the *residue of the double differential* W. We denote the residue ξ of W by $\mathfrak{R}(W)$, i. e. we put

$$\mathfrak{R}(W) = R_p W_{12}(\phi_p)\sigma_p. \tag{3.14}$$

As (3.12) shows, *the residue* $\xi = \mathfrak{R}(W)$ *is characterized by the equation*

$$dW[X] = 2\pi i \int_{\Gamma} \xi \cdot X, \tag{3.15}$$

where X is a "variable" 2-form of class C^{∞} with $|X| \subset \mathfrak{D}$. The above Theorem 3.2 says that *the residue* ξ *of any double differential defined in* $\mathfrak{U}_{\mathfrak{p}}$ *satisfies* $\sum_{\phi(p)=\mathfrak{p}} \mathrm{Res}_p[\xi] = 0$.

The differential σ_p depends on the choice of R_p and the local coordinates z^1, z^2 in the neighborhood $\mathfrak{U}_{\mathfrak{p}}$ of $\mathfrak{p} = \phi(p)$. Denoting the local coordinates in $\mathfrak{U}_{\phi(p)}$ by z_p^1, z_p^2, we get, from (3.13),

$$\sigma_q = [\partial(z_q^1, z_q^2)/\partial(z_p^1, z_p^2)]_{z=\phi(\tau)} t_{pq}\sigma_p, \qquad \text{in } U_p \cap U_q, \tag{3.16}$$

where $t_{pq} = [R_p(z)/R_q(z)]_{z=\phi(\tau)}$. Combined with (2.3), this yields immediately

$$h_q\sigma_q = [\partial(z_q^1, z_q^2)/\partial(z_p^1, z_p^2)]_{z=\phi(\tau)} h_p\sigma_p, \qquad \text{in } U_p \cap U_q. \tag{3.17}$$

For an arbitrary *meromorphic* differential $\xi = f d\tau_p$ defined in a neighborhood U_p of p, we put $V_p(\xi) = V_p(f)$. By the *conductor* $\mathfrak{c}$ of Γ will be meant the divisor $\mathfrak{c}$ defined as

$$\mathfrak{c} = \sum_p c_p p, \qquad c_p = - V_p(\sigma_p).$$

It is to be noted here that $V_p(\sigma_p)$ is independent of the choice of R_p and z^1, z^2, whereas σ_p depends on them (see (3.16)). Now we shall prove that

$$\begin{cases} c_p = 0, & \textit{if } \mathfrak{p} = \phi(p) \textit{ is a simple point,} \\ c_p > 0, & \textit{if } \mathfrak{p} = \phi(p) \textit{ is a singular point.} \end{cases}$$

To do this, we denote the system of local coordinates with the origin $\mathfrak{p}$ by (w, z). In case $\mathfrak{p} = \phi(p)$ is a simple point, we may assume that $R_{\mathfrak{p}}(w, z) = w$ and that $z = \tau_p$ on Γ. Hence we get $\sigma_p = d\tau_p$, proving that $c_p = 0$. In case $\mathfrak{p} = \phi(p)$ is a singular point, $\phi^{-1}(\mathfrak{p})$ consists of a finite number of points $p_1 = p, p_2, \cdots, p_l$. Let $R_{\mathfrak{p}}(w, z) = \prod_{s=1}^{l} R_s(w, z)$ be the decomposition of $R_{\mathfrak{p}}(w, z)$ into a product of the irreducible distinguished polynomials $R_s(w, z)$

in w representing respectively the irreducible branches Γ_s of Γ corresponding to p_s (see (3.2)). Then, introducing the parametric representation

$$\begin{cases} w = \mathfrak{P}_1(\tau_1) = \tau_1^{k_1}(a_0 + a_1\tau_1 + \cdots), & (a_0 \neq 0, k_1 \geqq 1), \\ z = \tau_1^{m_1} \end{cases}$$

of the branch $\Gamma_1 = \Gamma_p$, we have $d\phi_p^2 = m_1\tau_1^{m_1-1}d\tau_1$, while, since

$$\partial_1 R_{\mathfrak{p}}(\mathfrak{P}_1, \tau_1^{m_1})) = \partial_1 R_1(\mathfrak{P}_1, \tau_1^{m_1}) \cdot \prod_{s=2}^{l} R_s(\mathfrak{P}_1, \tau_1^{m_1}),$$

$V_p(\partial_1 R_{\mathfrak{p}}(\mathfrak{P}_1, \tau_1^{m_1}))$ is greater than $m_1 - 1$ unless $l = 1$ and $k_1 = 1$. Hence

$$c_p = V_p(\partial_1 R_{\mathfrak{p}}(\mathfrak{P}_1, \tau_1^{m_1})) - m_1 + 1 > 0$$

if $\mathfrak{p} = \phi(p)$ is a singular point, q. e. d. The above result shows that *the carrier* $|\mathfrak{c}|$ *of the conductor* $\mathfrak{c}$ *coincides with the inverse image* $\phi^{-1}(\mathfrak{S})$ *of the set* $\mathfrak{S}$ *consisting of all singular points of* Γ.

Now, let us consider the space $\mathfrak{F}(\Gamma)$. Let F be an arbitrary function in $\mathfrak{F}(\Gamma)$. Then R_pF is a holomorphic function in a neighborhood $\mathfrak{U}_{\mathfrak{p}}$ of $\mathfrak{p} = \phi(p)$ and therefore $f(\tau) = R_pF(\phi_p(\tau))/h_p(\tau)$ is a meromorphic function of τ in U_p, where $R_pF(\phi_p(\tau))$ denotes the holomorphic function of τ obtained from the holomorphic function R_pF of z^1, z^2 by putting $z^\alpha = \phi_p^\alpha(\tau)$ $(\alpha = 1, 2)$. Moreover, since by (2.12), $R_pF(\phi_p)/R_qF(\phi_q) = h_p/h_q$, in $U_p \cap U_q$, we have $R_pF(\phi_p)/h_p = R_qF(\phi_q)/h_q$, in $U_p \cap U_q$. This shows that $f = R_pF(\phi_p)/h_p$ *is a well defined meromorphic function on* $\mathbf{\Gamma}$. *We denote this function* f *by* $\mathfrak{S}_h(F)$, i. e. *we put* $\mathfrak{S}_h(F) = R_pF(\phi_p)/h_p$. The linear operator $\mathfrak{S}_h$ thus defined depends on the choice of the system $\{h_p\}$ (the subscript h in $\mathfrak{S}_h$ indicates this fact). Since $fh_p = R_pF(\phi_p)$ is holomorphic in U_p, the function $f = \mathfrak{S}_h(F)$ is a multiple of $-\mathfrak{d}$, i. e.

$$(\alpha) \qquad (f) + \mathfrak{d} \geqq 0.$$

Again, at each singular point $\mathfrak{p}$ of Γ, f satisfies

$$(\beta)^* \qquad \sum_{\phi(p)=\mathfrak{p}} \mathrm{Res}_p[B_{12}(\phi_p) f h_p \sigma_p] = 0$$

for an arbitrary holomorphic double differential $B_{12}dz^1dz^2$ defined in some (not fixed) neighborhood of $\mathfrak{p}$; indeed, we get, by Theorem 3.2,

$$\sum_{\phi(p)=\mathfrak{p}} \mathrm{Res}_p[B_{12} f h_p \sigma_p] = \sum_{\phi(p)=\mathfrak{p}} \mathrm{Res}_p[B_{12} R_{\mathfrak{p}} F \sigma_p] = 0,$$

since $R_{\mathfrak{p}}F$ is holomorphic. The above condition $(\beta)^*$ is obviously equivalent to the following condition

$$(\beta) \qquad \sum_{\phi(p)=\mathfrak{p}} \mathrm{Res}_p[(\phi_p^1)^k (\phi_p^2)^l f h_p \sigma_p] = 0, \qquad (k, l = 0, 1, 2, \cdots).$$

We denote by $\mathfrak{f}(\Gamma, \mathfrak{d})$ *the linear space consisting of all meromorphic functions* f *on* $\mathbf{\Gamma}$ *satisfying the above conditions* (α) *and* (β). Then the above results show that $F \rightarrow f = \mathfrak{S}_h(F)$ *is a linear transformation mapping* $\mathfrak{F}(\Gamma)$ *into* $\mathfrak{f}(\Gamma, \mathfrak{d})$; *moreover,* F *is determined uniquely up to an additive constant by* $f = \mathfrak{S}_h(F)$. Indeed, if $f = \mathfrak{S}_h(F) = R_pF/h_p$ vanishes identically, then the holomorphic function R_pF must be divisible by R_p and therefore F is holomorphic everywhere on $\mathfrak{M}$; hence F must be a constant.

Now, suppose an arbitrary meromorphic function $f \varepsilon \mathfrak{f}(\Gamma, \mathfrak{d})$ as given. Then, does there exist a meromorphic function $F \varepsilon \mathfrak{F}(\Gamma)$ such that $\mathfrak{S}_h(F) = f$? In order to solve this question, it is convenient to introduce the notion of *additive* meromorphic functions. A many valued meromorphic function $F(z)$ on $\mathfrak{M}$ will be called *additive,* if, by the analytic continuation along an arbitrary closed continuous curve γ on $\mathfrak{M}$, $F(z)$ is changed into $F(z) + c_\gamma$, where c_γ is a constant depending on γ. An arbitrary additive meromorphic function is represented as the integral $\int^z dF$ of the *simple differential* dF; thus an additive meromorphic function is a simple Picard integral of the first or the second kind according as it is everywhere holomorphic or it is not. Let $\mathfrak{A}(\Gamma)$ be the linear space consisting of all *additive* meromorphic functions F on $\mathfrak{M}$ which are multiples of $-\Gamma$ in the sense that, for every point $\mathfrak{p}$ on $\mathfrak{M}$, $R_\mathfrak{p}F$ is holomorphic in some neighborhood of $\mathfrak{p}$. Then it can be shown in the same way as above that, for each $F \varepsilon \mathfrak{A}(\Gamma)$, $f = \mathfrak{S}_h(F) = R_pF/h_p$ is a (single valued) meromorphic function belonging to $\mathfrak{f}(\Gamma, \mathfrak{d})$. Furthermore it follows from (3.7) and (3.9) that $f = \mathfrak{S}_h(F)$ is related with F by the relation

$$(1 + iC)\,dF[\Psi] = 4\pi i \int_{\mathbf{\Gamma}} \sum_{\beta} \Psi_{12\beta\bullet}(\phi_p) f h_p \sigma_p d\bar{\phi}_p{}^{\beta}, \tag{3.18}$$

where Ψ is an arbitrary 3-form of class C^∞ and the value of the integral in the right-hand side is to be considered as the Cauchy principal value if it does not converge absolutely. Now, we shall prove that, *for every* $f \varepsilon \mathfrak{f}(\Gamma, \mathfrak{d})$, *there exists at least one additive meromorphic function* $F \varepsilon \mathfrak{A}(\Gamma)$ *such that* $\mathfrak{S}_h(F) = f$. For this purpose, we introduce, in view of (3.18), the 1-current Q_f defined as

$$Q_f[\Psi] = -\int_{\mathbf{\Gamma}} \sum_{\beta} \Psi_{12\beta\bullet}(\phi_p) f h_p \sigma_p d\bar{\phi}_p{}^{\beta}, \tag{3.19}$$

where Ψ is a "variable" 3-form of class C^∞ and the value of the integral in the right-hand side is to be considered as the Cauchy principal value in case it does not converge absolutely. It is to be noted here that we have, by virtue of (3.17),

$$\sum_\beta \Psi_{12\beta^*}(\phi_p) f h_p \sigma_p d\bar{\phi}_p{}^\beta = \sum_\beta \Psi_{12\beta^*}(\phi_q) f h_q \sigma_q d\bar{\phi}_q{}^\beta, \qquad \text{in } U_p \cap U_q,$$

so that $\sum_\beta \Psi_{12\beta^*}(\phi_p) f h_p \sigma_p d\bar{\phi}_p{}^\beta$ is a well defined 2-form on $\mathbf{\Gamma}$.

It is obvious that *the carrier* $|Q_f|$ *of the* 1*-current* Q_f is Γ. Again Q_f *is a pure* 1*-current of type* 1, as (3.19) shows. Now we shall prove that dQ_f satisfies

$$CdQ_f = dQ_f. \tag{3.20}$$

To do this, let Φ be an arbitrary 2-form of class C^∞. Then we have

$$dQ_f[\Phi] = Q_f[d\Phi] = -\int_{\mathbf{\Gamma}} \sum_\beta (\partial_1 \Phi_{2\beta^*} - \partial_2 \Phi_{1\beta^*} + \partial_{\beta^*}\Phi_{12}) f h_p \sigma_p d\bar{\phi}_p{}^\beta.$$

Considering $\Phi_{12}(\phi_p) f h_p \sigma_p$ as a 1-form on $\mathbf{\Gamma}$, we get

$$d(\Phi_{12}(\phi_p) f h_p \sigma_p) = -\sum_\beta \partial_{\beta^*}\Phi_{12} f h_p \sigma_p d\bar{\phi}_p{}^\beta$$

and consequently

$$\int_{\mathbf{\Gamma}} \sum_\beta \partial_{\beta^*}\Phi_{12} f h_p \sigma_p d\bar{\phi}_p{}^\beta = -\int_{\mathbf{\Gamma}} d(\Phi_{12} f h_p \sigma_p) = \sum_{\mathfrak{p}} \sum_{\phi(p)=\mathfrak{p}} \lim_{\epsilon \to +0} \int_{|\tau_p|=\epsilon} \Phi_{12} f h_p \sigma_p,$$

where the sum $\sum_{\mathfrak{p}}$ is to be extended over all singular points of Γ; while, as Lemma 3.1 shows, the condition $(\beta)^*$ implies that

$$\sum_{\phi(p)=\mathfrak{p}} \lim_{\epsilon \to +0} \int_{|\tau_p|=\epsilon} \Phi_{12} f h_p \sigma_p = 0.$$

Hence we obtain

$$dQ_f[\Phi] = -\int_{\mathbf{\Gamma}} \sum_\beta (\partial_1 \Phi_{2\beta^*} - \partial_2 \Phi_{1\beta^*}) f h_p \sigma_p d\bar{\phi}_p{}^\beta,$$

proving that $dQ_f[C\Phi] = dQ_f[\Phi]$. Combined with (1.6), this yields immediately (3.20). Thus dQ_f *is a pure* 2*-current of type* 1.

Now, we put $\theta_f = (d\Lambda G + i\delta G) dQ_f$. Then, by virtue of Theorem 1.4, θ_f is a holomorphic simple differential in $\mathfrak{M} - \Gamma$ and the integral of θ_f over an arbitrary 1-cycle $\gamma \subset \mathfrak{M} - \Gamma$ is given by

$$\int_\gamma \theta_f = iQ_f[H\gamma]. \tag{3.21}$$

Putting $F_f(z) = 2\pi \int^z \theta_f$, we get therefore an additive holomorphic function

F_f in $\mathfrak{M}-\Gamma$. Now we shall show that, for every $\mathfrak{p}\,\varepsilon\,\Gamma$, $R_\mathfrak{p}F_f$ is holomorphic in $\mathfrak{U}_\mathfrak{p}$. First, assume that $\mathfrak{p}$ is a simple point of Γ and put $p=\phi^{-1}(\mathfrak{p})$. Then we can choose the system of local coordinates (w,z) with the origin $\mathfrak{p}$ so that $R_\mathfrak{p}(w,z)=w$ and that $z=\tau_p$ on Γ. The holomorphic function $f(\tau)h_p(\tau)$ defined on $\Gamma\cap\mathfrak{U}_\mathfrak{p}$ can be extended to a holomorphic function $N(w,z)$ defined in $\mathfrak{U}_\mathfrak{p}$, e. g. we get such $N(w,z)$ by putting simply $N(w,z)=f(z)h_p(z)$. Using (3.7) and (3.9), we infer immediately that the 0-current $N/R_\mathfrak{p}$ in $\mathfrak{U}_\mathfrak{p}$ satisfies

$$(1+iC)d(N/R_\mathfrak{p})=-4\pi iQ_f. \tag{3.22}$$

Now, consider the difference

$$T=2\pi\theta_f-\tfrac{1}{2}(1-iC)d(N/R_\mathfrak{p}). \tag{3.23}$$

This difference T is a pure 1-current of type 0 in $\mathfrak{U}_\mathfrak{p}$ and, since, by Theorem 1.4, θ_f satisfies $d\theta_f=idQ_f$, it follows from (3.22) that T satisfies $dT=0$. Hence, by Theorem 1.3, T is a holomorphic differential in $\mathfrak{U}_\mathfrak{p}$. The integral

$$L(w,z)=\int^{(w,z)}T$$

is therefore a holomorphic function in $\mathfrak{U}_\mathfrak{p}$. On the other hand, we have

$$\tfrac{1}{2}(1-iC)d(N/R_\mathfrak{p})=d(N/R_\mathfrak{p}),\qquad \text{in } \mathfrak{U}_\mathfrak{p}-\Gamma.$$

Hence we get, from (3.23),

$$F_f=N/R_\mathfrak{p}+L+\text{const.},\qquad \text{in } \mathfrak{U}_\mathfrak{p}-\Gamma. \tag{3.24}$$

This shows that $R_\mathfrak{p}F_f$ is a holomorphic function in $\mathfrak{U}_\mathfrak{p}$. Moreover (3.24) yields immediately $R_\mathfrak{p}F(0,z)=N(0,z)=f(z)h_p(z)$, proving that

$$R_pF(\phi_p)/h_p=f,\quad \text{in } \mathfrak{U}_\mathfrak{p}. \tag{3.25}$$

Now, consider the case that $\mathfrak{p}$ is a singular point of Γ. Then it follows from the above result that $R_\mathfrak{p}F_f$ is holomorphic in $\mathfrak{U}_\mathfrak{p}-\mathfrak{p}$. Hence, by a theorem of Hartogs,[37] $R_\mathfrak{p}F_f$ must be holomorphic in the whole of $\mathfrak{U}_\mathfrak{p}$. Thus F_f is a function belonging to $\mathfrak{A}(\Gamma)$. Furthermore it follows from (3.25) that F_f satisfies $\mathfrak{S}_h(F_f)=f$. Obviously every $F\,\varepsilon\,\mathfrak{A}(\Gamma)$ satisfying $\mathfrak{S}_h(F)=f$ can be represented by means of the base $\{A_1,A_2,\cdots,A_q\}$ of simple differentials of the first kind as

$$F(z)=F_f(z)+\int^z\sum_{\nu=1}^{q}\eta_\nu A_\nu+\text{const.},$$

since $F-F_f$ is holomorphic everywhere in $\mathfrak{M}$. Thus we obtain the following

---

[37] Hartogs [11].

THEOREM 3. 3. *The linear mapping* $F \rightarrow f = \mathfrak{S}_h(F)$ *maps* $\mathfrak{A}(\Gamma)$ *onto* $\mathfrak{f}(\Gamma, \mathfrak{d})$. *Every* $F \varepsilon \mathfrak{A}(\Gamma)$ *can be written by means of* $f = \mathfrak{S}_h(F)$ *as*

$$F(z) = \int^{z} (2\pi\theta_f + \sum_{\nu=1}^{q} \eta_\nu A_\nu) + \eta_0, \tag{3.26}$$

where $\eta_0, \eta_1, \eta_2, \cdots, \eta_q$ *are constants. Conversely, for arbitrary given* $\eta_0, \eta_1, \eta_2, \cdots, \eta_q$, *the function* F *defined by* (3. 26) *belongs to* $\mathfrak{A}(\Gamma)$ *and satisfies* $\mathfrak{S}_h(F) = f$.

It is to be noted here that the additive meromorphic function F defined by (3. 26) satisfies as a 0-current the relation

$$dF = 2\pi\theta_f + \sum \eta_\nu A_\nu - 2\pi i Q_f. \tag{3.27}$$

Indeed, putting $T = dF - 2\pi\theta_f - \sum \eta_\nu A_\nu + 2\pi i Q_f$, we infer readily that dT vanishes in the whole of $\mathfrak{M}$ and that T vanishes in $\mathfrak{M} - \Gamma$, while, since, by (3. 18), $(1 + iC) dF = -4\pi i Q_f$, we get $(1 + iC) T = 0$, showing that T is a pure 1-current of type 0. Hence, by virtue of Theorem 1. 3, T is a holomorphic differential and therefore T must vanish in the whole of $\mathfrak{M}$.

Let $\{\gamma_1, \gamma_2, \cdots, \gamma_{2q}\}$ be a base of 1-cycles on $\mathfrak{M}$ with integral coefficients and $\{A_1, \cdots, A_q\}$ be a base of simple differentials of the first kind such that $(A_\lambda, \bar{A}_\nu) = \delta_{\lambda\nu}$. Again we put

$$\omega_{\nu j} = \gamma_j[A_\nu] = \int_{\gamma_j} A_\nu.$$

$(\omega_{\nu j})$ is called the *period matrix* of the simple differentials of the first kind. Since $HA_\nu = A_\nu$, we have $\omega_{\nu j} = (A_\nu, {*H}\gamma_j)$. Obviously A_λ can be represented as

$$A_\lambda = \sum_{j=1}^{2q} \alpha_{\lambda j} {*H}\gamma_j, \tag{3.28}$$

where the coefficients $\alpha_{\lambda j}$ are related with $\omega_{\nu j}$ by the formulae

$$\sum_j \alpha_{\lambda j}\omega_{\nu j} = 0, \qquad \sum_j \bar{\alpha}_{\lambda j}\omega_{\nu j} = \delta_{\lambda\nu}. \tag{3.29}$$

Using (3. 21), we infer that the period $\omega_j(F) = \int_{\gamma_j} dF$ of the additive meromorphic function F defined by (3. 26) is given by

$$\omega_j(F) = 2\pi i Q_f[H\gamma_j] + \sum_\nu \eta_\nu \omega_{\nu j}. \tag{3.30}$$

Putting

$$\alpha_\lambda(F) = \sum_j \alpha_{\lambda j}\omega_j(F), \qquad \alpha_{\lambda^*}(F) = \sum_j \bar{\alpha}_{\lambda j}\omega_j(F),$$

we introduce the λ- and the λ^*- components $\alpha_\lambda(F)$, $\alpha_{\lambda^*}(F)$ of the "period vector" $\omega_j(F)$ of F. By means of the "fundamental form" $\Omega = \sum i g_{\alpha\beta^*} dz^\alpha d\bar{z}^\beta$, $*A_\lambda$ can be represented as[38] $*A_\lambda = -i\Omega\cdot A_\lambda$. Hence we have, by (3.28),

$$\sum_{j=1}^{2q} \alpha_{\lambda j} H\gamma_j = i\Omega\cdot A_\lambda.$$

Using this formula, (3.29) and the identity $Q_f[\Omega\cdot\bar{A}_\lambda]=0$, we get from (3.30)

$$\alpha_\lambda(F) = -2\pi Q_f[\Omega\cdot A_\lambda], \qquad \alpha_{\lambda^*}(F) = \eta_\lambda. \tag{3.31}$$

THEOREM 3.4 (EXISTENCE THEOREM). *Let f be an arbitrary function in $\mathfrak{f}(\Gamma,\mathfrak{d})$. Then, there exists on $\mathfrak{M}$ a function $F\,\varepsilon\,\mathfrak{F}(\Gamma)$ such that $\mathfrak{S}_h(F)=f$ if and only if f satisfies*

$$Q_f[\Omega\cdot A_\lambda]=0, \qquad \text{for } \lambda=1,2,\cdots,q, \tag{3.32}$$

and, in case f satisfies (3.32), *F is given by*

$$F(z) = 2\pi\int^z \theta_f + \text{const.} \tag{3.33}$$

Proof. Obviously $F\,\varepsilon\,\mathfrak{A}(\Gamma)$ belongs to $\mathfrak{F}(\Gamma)$ if and only if $\alpha_\lambda(F) = \alpha_{\lambda^*}(F)=0$ $(\lambda=1,2,\cdots,q)$. Hence, if F belongs to $\mathfrak{F}(\Gamma)$, then, by (3.31), $f=\mathfrak{S}_h(F)$ must satisfy (3.32). Conversely, if f satisfies (3.32), then, putting $\eta_\lambda=0$ in (3.26), we get $F\,\varepsilon\,\mathfrak{F}(\Gamma)$ satisfying $\mathfrak{S}_h(F)=f$, q. e. d.

Since $F\,\varepsilon\,\mathfrak{F}(\Gamma)$ is determined by $f=\mathfrak{S}_h(F)$ uniquely up to an additive constant, $\dim\mathfrak{F}(\Gamma)-1$ is equal to the number of linearly independent functions f in $\mathfrak{f}(\Gamma,\mathfrak{d})$ satisfying (3.32). Denote by j the number of linearly independent simple differentials A of the first kind satisfying $Q_f[\Omega\cdot A]=0$, for all $f\,\varepsilon\,\mathfrak{f}(\Gamma,\mathfrak{d})$. Then (3.32) imposes $q-j$ linearly independent conditions on f. Hence we get

$$\dim\{\Gamma\} = \dim\mathfrak{f}(\Gamma,\mathfrak{d}) - q + j + 1. \tag{3.34}$$

The non-negative integer $\dim\mathfrak{f}(\Gamma,\mathfrak{d}) - \dim\{\Gamma\}+1$ will be called the *characteristic deficiency*[39] of Γ and be denoted by $chd(\Gamma)$:

$$chd(\Gamma) = \dim\mathfrak{f}(\Gamma,\mathfrak{d}) - \dim\{\Gamma\}+1.$$

Then, it follows immediately from (3.34) that

$$chd(\Gamma) = q - j; \tag{3.35}$$

---

[38] Hodge [12], p. 175.

[39] Goldman [9], § 6. The characteristic deficiency is nothing other than *the deficiency of the characteristic series* of Italian geometers. Cf. Zariski [30], p. 26, pp. 69-71.

thus we see that *the characteristic deficiency of an arbitrary curve is always not greater than* q.

A finite number of simple Picard integrals of the second kind are said to be *independent* if no non-trival linear combination of these integrals reduces to a single valued meromorphic function. By virtue of Theorem 3. 3, we have $\dim \mathfrak{A}(\Gamma) = \dim \mathfrak{f}(\Gamma, \mathfrak{d}) + q + 1$, while the number of independent simple Picard integrals of the second kind which are multiples of $-\Gamma$ is obviously equal to $\dim \mathfrak{A}(\Gamma) - \dim \mathfrak{F}(\Gamma)$. Hence we obtain

THEOREM 3. 5. *The number of independent simple Picard integrals of the second kind which are multiples of* $-\Gamma$ *is equal to* $q + chd(\Gamma)$.

It follows from this theorem that, *if* $\mathfrak{M}$ *contains a curve with the characteristic deficiency* q, $\mathfrak{M}$ *possesses exactly* $2q$ *independent simple Picard integrals of the second kind.*

4. The theorem of Riemann-Roch on reducible curves. By virtue of the formula (3. 34), our problem is reduced to the estimation of $\dim \mathfrak{f}(\Gamma, \mathfrak{d})$. Let Σ be an irreducible curve without singular point. For an arbitrary differential ξ on Σ, we denote by (ξ) the divisor of ξ defined as $(\xi) = \sum_p V_p(\xi) \cdot p$. Now, for an arbitrary divisor $\mathfrak{m}$ on Σ, we denote by $\mathfrak{f}(\mathfrak{m})$ the space of all meromorphic functions on Σ satisfying $(f) + \mathfrak{m} \geqq 0$ and put $\mathfrak{f}^*(\mathfrak{m}) = \mathfrak{f}(\mathfrak{m}) + \mathfrak{f}(0)$ ($\mathfrak{f}(0)$ is the space consisting of all constants). In case $\mathfrak{m} \geqq 0$, $\mathfrak{f}^*(\mathfrak{m})$ coincides with $\mathfrak{f}(\mathfrak{m})$. Again, we denote by $\Xi(\mathfrak{m})$ the space of all differentials ξ on Σ satisfying $(\xi) + \mathfrak{m} \geqq 0$ and put $\Xi^*(\mathfrak{m}) = \Xi(\mathfrak{m}) + \Xi(0)$. $\Xi(0)$ is the space of all differentials of the first kind and therefore $\dim \Xi(0)$ is equal to the genus π of Σ.

Let $\mathfrak{a} = \sum m_p p$, $\mathfrak{b} = \sum n_q q$ be two *positive* divisors on Σ such that the carriers $|\mathfrak{a}|$ and $|\mathfrak{b}|$ have no point in common. Now, for arbitrary $\xi \varepsilon \Xi(\mathfrak{b})$ and $f \varepsilon \mathfrak{f}(\mathfrak{a})$, we define the product (ξf) as

$$(\xi f) = \sum_{q \varepsilon |\mathfrak{b}|} \mathrm{Res}_q[f\xi]\,; \text{ hence } (\xi f) = - \sum_{p \varepsilon |\mathfrak{a}|} \mathrm{Res}_p[f\xi].$$

As one readily infers, (ξf) *vanishes for all* $\xi \varepsilon \Xi(\mathfrak{b})$ *if and only if* $f \varepsilon \mathfrak{f}^*(\mathfrak{a} - \mathfrak{b})$. Again, if $\xi \varepsilon \Xi^*(\mathfrak{b} - \mathfrak{a})$, *then* (ξf) *vanishes for all* $f \varepsilon \mathfrak{f}(\mathfrak{a})$. Thus, in the product (ξf), ξ and f can be considered as the elements of the factor spaces $\bar{\Xi} = \Xi(\mathfrak{b})/\Xi^*(\mathfrak{b} - \mathfrak{a})$ and $\bar{\mathfrak{f}} = \mathfrak{f}(\mathfrak{a})/\mathfrak{f}^*(\mathfrak{a} - \mathfrak{b})$, respectively. Now we shall show that ξ *belongs to* $\Xi^*(\mathfrak{b} - \mathfrak{a})$ *if* (ξf) *vanishes for all* $f \varepsilon \mathfrak{f}(\mathfrak{a})$ so that $\bar{\Xi}$ *and* $\bar{\mathfrak{f}}$ *are dual to each other with respect to the product* (ξf). To prove this, it is sufficient to verify that $\dim \bar{\Xi} = \dim \bar{\mathfrak{f}}$. By virtue of the theorem of Riemann-Roch, we have

$$\dim \mathfrak{f}(\mathfrak{a}) = \deg \mathfrak{a} - \pi + i(\mathfrak{a}) + 1,$$
$$\dim \mathfrak{f}^*(\mathfrak{a} - \mathfrak{b}) = \deg \mathfrak{a} - \deg \mathfrak{b} - \pi + i(\mathfrak{a} - \mathfrak{b}) + 2,$$

where π is the genus of Σ and $i(\mathfrak{a})$, $i(\mathfrak{a} - \mathfrak{b})$ are the indices of specialties of of $\mathfrak{a}$, $\mathfrak{a} - \mathfrak{b}$, respectively. Hence we get

$$\dim \bar{\mathfrak{f}} = \deg \mathfrak{b} - 1 + i(\mathfrak{a}) - i(\mathfrak{a} - \mathfrak{b}).$$

On the other hand, since

$$i(\mathfrak{a}) = \dim \Xi(-\mathfrak{a}) = \dim [\Xi(\mathfrak{b} - \mathfrak{a}) \cap \Xi(0)],\ i(\mathfrak{a} - \mathfrak{b}) = \dim \Xi(\mathfrak{b} - \mathfrak{a}),$$

we have

$$\dim \Xi^*(\mathfrak{b} - \mathfrak{a}) = i(\mathfrak{a} - \mathfrak{b}) + \pi - i(\mathfrak{a}),$$

whence we get, using $\dim \Xi(\mathfrak{b}) = \deg \mathfrak{b} - 1 + \pi$,

$$\dim \bar{\Xi} = \deg \mathfrak{b} - 1 + i(\mathfrak{a}) - i(\mathfrak{a} - \mathfrak{b}).$$

Hence we obtain $\dim \bar{\Xi} = \dim \bar{\mathfrak{f}}$, q. e. d.

Now we turn to the reducible curve $\Gamma = \sum_{\nu=1}^{\mu} \Gamma_\nu$ on $\mathfrak{M}$. For an arbitrary divisor $\mathfrak{m}$ on the non-singular model $\mathbf{\Gamma}$ of Γ, we denote by $\mathfrak{f}(\mathfrak{m})$ the space of all meromorphic functions f on $\mathbf{\Gamma}$ satisfying $(f) + \mathfrak{m} \geqq 0$, by $\Xi(\mathfrak{m})$ the space of all differentials ξ on $\mathbf{\Gamma}$ satisfying $(\xi) + \mathfrak{m} \geqq 0$ and put $\mathfrak{f}^*(\mathfrak{m}) = \mathfrak{f}(\mathfrak{m}) + \mathfrak{f}(0)$, $\Xi^*(\mathfrak{m}) = \Xi(\mathfrak{m}) + \Xi(0)$. We choose the characteristic divisor $\mathfrak{d}$ of Γ so that $|\mathfrak{d}|$ and $|\mathfrak{c}|$ have no point in common and decompose $\mathfrak{d}$ as follows:

$$\mathfrak{d} = \mathfrak{d}^+ - \mathfrak{d}^-, \quad \mathfrak{d}^+ \geqq 0,\ \mathfrak{d}^- \geqq 0 \text{ and } |\mathfrak{d}^+| \cap |\mathfrak{d}^-| \text{ is empty.}$$

Again, we put $\mathfrak{a} = \mathfrak{d}^+$, $\mathfrak{b} = \mathfrak{d}^- + \mathfrak{c}$ and introduce the product

$$(\xi f) = \sum_{q \varepsilon |\mathfrak{b}|} \operatorname{Res}_q [f\xi]$$

of arbitrary $\xi \varepsilon \Xi(\mathfrak{b})$ and $f \varepsilon \mathfrak{f}(\mathfrak{a})$. The non-singular model $\mathbf{\Gamma}$ is the direct sum: $\mathbf{\Gamma} = \sum_\nu \mathbf{\Gamma}_\nu$ of irreducible components $\mathbf{\Gamma}_\nu$. Corresponding to this, $\mathfrak{a}$, $\mathfrak{b}$, ξ, f can be decomposed into the direct sums: $\mathfrak{a} = \sum_\nu \mathfrak{a}_\nu$, $\mathfrak{b} = \sum_\nu \mathfrak{b}_\nu$, $\xi = \sum_\nu \xi_\nu$, $f = \sum_\nu f_\nu$ of $\mathfrak{a}_\nu$, $\mathfrak{b}_\nu$, ξ_ν, f_ν defined on $\mathbf{\Gamma}_\nu$ so that we have

$$(\xi f) = \sum_\nu (\xi_\nu f_\nu), \qquad \xi_\nu \varepsilon \Xi(\mathfrak{b}_\nu),\ f_\nu \varepsilon \mathfrak{f}(\mathfrak{a}_\nu),$$

where $(\xi_\nu f_\nu) = \sum_{q \varepsilon |\mathfrak{b}_\nu|} \operatorname{Res}_q [f_\nu \xi_\nu]$. Hence, applying the "duality theorem" proved above to each pair of divisors $\mathfrak{a}_\nu$, $\mathfrak{b}_\nu$, we see immediately that the *factor spaces* $\Xi(\mathfrak{d}^- + \mathfrak{c})/\Xi^*(\mathfrak{c} - \mathfrak{d})$ and $\mathfrak{f}(\mathfrak{d}^+)/\mathfrak{f}^*(\mathfrak{d} - \mathfrak{c})$ *are dual to each other with respect to the product* (ξf).

It is convenient for our present purpose to introduce the following notations: Let $\mathfrak{m} = \sum m_p p$ be a divisor on $\mathbf{\Gamma}$ and $\{\xi_p \mid p \in \mathfrak{T}\}$ be a system of differentials ξ_p defined respectively in a neighborhood U_p of $p \in \mathfrak{T}$, where $\mathfrak{T}$ is a finite set of points. Again, let ξ be a differential on $\mathbf{\Gamma}$. Then, we shall say that ξ is congruent to $\sum_{p \in \mathfrak{T}} \xi_p$ modulo $\mathfrak{m}$ and write

$$\xi \equiv \sum_{p \in \mathfrak{T}} \xi_p \quad (\mathfrak{m}), \tag{4.1}$$

if ξ satisfies

$$V_p(\xi - \xi_p) \geqq m_p, \text{ for } p \in \mathfrak{T}; \qquad V_p(\xi) \geqq m_p, \text{ for } p \notin \mathfrak{T}.$$

The *formal sum* $\sum \xi_p$ in the right-hand side of (4. 1) stands simply for the *system* or the *vector* $\{\xi_p\}$. Again, a finite number of formal sums $\sum_p \xi_p^{(\nu)}$ $(\nu = 1, 2, \cdots)$ will be called *independent modulo* $\mathfrak{m}$, if no non-trivial linear combination $\sum a_\nu \sum_p \xi_p^{(\nu)}$ satisfies

$$0 \equiv \sum a_\nu \sum_p \xi_p^{(\nu)} \quad (\mathfrak{m}).$$

For each singular point $\mathfrak{p}$ of Γ, we introduce the formal sums

$$\sigma_{\mathfrak{p}}^{(k,l)} = \sum_{\phi(p)=\mathfrak{p}} (\phi_p^1)^k (\phi_p^2)^l \sigma_p, \qquad (k, l = 0, 1, 2, \cdots)$$

and denote by $l_{\mathfrak{p}}$ the number of $\sigma_{\mathfrak{p}}^{(k,l)}$ which are independent modulo 0, where 0 denotes the divisor *zero*: $0 = \sum_p 0 \cdot p$. If one considers

$$(\beta_{\mathfrak{p}}) \qquad \sum_{\phi(p)=\mathfrak{p}} \operatorname{Res}_p[(\phi_p^1)^k (\phi_p^2)^l f h_p \sigma_p] = 0 \qquad (k, l = 0, 1, 2, \cdots)$$

as the conditions for the system $\{f h_p \mid p \in \phi^{-1}(\mathfrak{p})\}$, then $l_{\mathfrak{p}}$ represents the number of linearly independent conditions involved in $(\beta_{\mathfrak{p}})$. For any simple point $\mathfrak{p}$, we put $l_{\mathfrak{p}} = 0$.

Now, let $\Xi(\Gamma, -\mathfrak{d})$ be the space consisting of all differentials ξ on $\mathbf{\Gamma}$ which are congruent modulo $\mathfrak{d}$ to some linear combinations of the formaal sums $\sum_{\phi(p)=\mathfrak{p}} (\phi_p^1)^k (\phi_p^2)^l h_p \sigma_p$, i. e.

$$\Xi(\Gamma, -\mathfrak{d}) = \{\xi \mid \xi \equiv \sum_{\mathfrak{p}} \sum_{k,l} a_{\mathfrak{p}kl} \sum_{\phi(p)=\mathfrak{p}} (\phi_p^1)^k (\phi_p^2)^l h_p \sigma_p \quad (\mathfrak{d})\},$$

where $a_{\mathfrak{p}kl}$ $(k, l = 0, 1, 2, \cdots)$ are arbitrary constants and $\sum_{\mathfrak{p}}$ is the sum extended over all singular points of Γ. Then we have

THEOREM 4. 1. *The dimension of* $\mathfrak{f}(\Gamma, \mathfrak{d})$ *is given by the formula*

$$\dim \mathfrak{f}(\Gamma, \mathfrak{d}) = \deg \mathfrak{d} - \sum_{\nu=1}^{\mu} \pi_\nu + \mu - \sum_{\mathfrak{p}} l_{\mathfrak{p}} + i, \tag{4.2}$$

where μ *is the number of irreducible components of* $\Gamma = \sum_{\nu=1}^{\mu} \Gamma_\nu$, π_ν *is the genus of* $\mathbf{\Gamma}_\nu$ *and* $i = \dim \Xi(\Gamma, -\mathfrak{d})$.

Proof. Assume first that $|\mathfrak{d}|$ and $|\mathfrak{c}|$ have no point in common and that each component $\mathfrak{d}^-_\nu$ of $\mathfrak{d}^-$ is *positive.* Then, for every $\sum_{\phi(p)=\mathfrak{p}} (\phi_p{}^1)^k (\phi_p{}^2)^l h_p \sigma_p$, there exists at least one differential ξ on $\mathbf{\Gamma}$ satisfying

$$\xi \equiv \sum_{\phi(p)=\mathfrak{p}} (\phi_p{}^1)^k (\phi_p{}^2)^l h_p \sigma_p \quad (-\mathfrak{d}^-).$$

Hence, denoting by Ξ the space consisting of all differentials ξ satisfying

$$\xi \equiv \sum_{\mathfrak{p}} \sum_{k,l} a_{\mathfrak{p}kl} \sum_{\phi(p)=\mathfrak{p}} (\phi_p{}^1)^k (\phi_p{}^2)^l h_p \sigma_p \quad (-\mathfrak{d}^-),$$

$a_{\mathfrak{p}kl}$ being arbitrary constants, we infer readily that the space $\mathfrak{f}^*(\Gamma, \mathfrak{d}) = \mathfrak{f}(\Gamma, \mathfrak{d}) + \mathfrak{f}(0)$ is the subspace of $\mathfrak{f}(\mathfrak{d}^+)$ consisting of all $f \in \mathfrak{f}(\mathfrak{d}^+)$ satisfying $(\xi f) = 0$, for all $\xi \in \Xi$; thus $\mathfrak{f}^*(\Gamma, \mathfrak{d})$ is characterized as the space which is orthogonal to Ξ in the dualism between $\mathfrak{f}(\mathfrak{d}^+)/\mathfrak{f}^*(\mathfrak{d} - \mathfrak{c})$ and $\Xi(\mathfrak{d}^- + \mathfrak{c})/\Xi^*(\mathfrak{c} - \mathfrak{d})$ with respect to the product (ξf). Obviously Ξ contains $\Xi(0)$ but does not necessarily contain $\Xi(\mathfrak{c} - \mathfrak{d})$. The space which is orthogonal to $\mathfrak{f}^*(\Gamma, \mathfrak{d})$ is therefore $\Xi + \Xi(\mathfrak{c} - \mathfrak{d})$. Hence we get

$$\dim \mathfrak{f}^*(\Gamma, \mathfrak{d}) - \dim \mathfrak{f}^*(\mathfrak{d} - \mathfrak{c}) = \dim \Xi(\mathfrak{d}^- + \mathfrak{c}) - \dim[\Xi + \Xi(\mathfrak{c} - \mathfrak{d})]. \tag{4.3}$$

Obviously we have

$$\dim \mathfrak{f}^*(\Gamma, \mathfrak{d}) = \dim \mathfrak{f}(\Gamma, \mathfrak{d}) + \mu,$$
$$\dim \mathfrak{f}^*(\mathfrak{d} - \mathfrak{c}) = \dim \mathfrak{f}(\mathfrak{d} - \mathfrak{c}) + \mu,$$
$$\dim \Xi(\mathfrak{d}^- + \mathfrak{c}) = \deg \mathfrak{c} + \dim \Xi(\mathfrak{d}^-)$$

and, since $\Xi \cap \Xi(\mathfrak{c} - \mathfrak{d}) = \Xi(\Gamma, -\mathfrak{d})$,

$$\dim[\Xi + \Xi(\mathfrak{c} - \mathfrak{d})] = \dim \Xi + \dim \Xi(\mathfrak{c} - \mathfrak{d}) - i,$$

while, applying the theorem of Riemann-Roch to each $\mathbf{\Gamma}_\nu$, we get

$$\dim \mathfrak{f}(\mathfrak{d} - \mathfrak{c}) = \deg \mathfrak{d} - \deg \mathfrak{c} - \sum \pi_\nu + \mu + \dim \Xi(\mathfrak{c} - \mathfrak{d}).$$

Inserting these formulae in (4.3) and using $\dim \Xi - \dim \Xi(\mathfrak{d}^-) = \sum_{\mathfrak{p}} l_{\mathfrak{p}}$, we obtain

$$\dim \mathfrak{f}(\Gamma, \mathfrak{d}) = \deg \mathfrak{d} - \sum \pi_\nu + \mu - \sum_{\mathfrak{p}} l_{\mathfrak{p}} + i.$$

Thus (4.2) is proved in the case that $\mathfrak{d}$ is subject to the conditions: $\mathfrak{d}^-_\nu > 0$,

$|\mathfrak{d}| \cap |\mathfrak{c}|$ is empty. Now, consider an arbitrary characteristic divisor $\mathfrak{d}'$ not satisfying the above conditions. $\mathfrak{d}'$ is determined by a system $\{h'_p\}$ which is obtained from $\{h_p\}$ by the transformation $h_p \rightarrow h'_p = g \cdot h_p$, g being a meromorphic function on $\mathbf{\Gamma}$. Then, as one readily infers, $\mathfrak{f}(\Gamma, \mathfrak{d}')$ or $\Xi(\Gamma, -\mathfrak{d}')$ is obtained respectively from $\mathfrak{f}(\Gamma, \mathfrak{d})$ or $\Xi(\Gamma, -\mathfrak{d})$ by the transformation $f \rightarrow f' = g^{-1} \cdot f$ or $\xi \rightarrow \xi' = g \cdot \xi$, so that we have

$$\dim \mathfrak{f}(\Gamma, \mathfrak{d}') = \dim \mathfrak{f}(\Gamma, \mathfrak{d}), \qquad \dim \Xi(\Gamma, -\mathfrak{d}') = \dim \Xi(\Gamma, -\mathfrak{d}).$$

Hence (4. 2) is valid also for $\mathfrak{d}'$, q. e. d.

5. The theorem of Riemann-Roch on compact analytic surfaces. Let $D = \sum m_\nu \Gamma_\nu$ be the representation of an arbitrary divisor D on $\mathfrak{M}$ in terms of the irreducible components Γ_ν of D. Then, at each point $\mathfrak{p} \varepsilon \mathfrak{M}$, D can be *represented* by the meromorphic function

$$R_{\mathfrak{p}}(D; z^1, z^2) = \prod_\nu \{R_{\nu\mathfrak{p}}(z^1, z^2)\}^{m_\nu},$$

where each $R_{\nu\mathfrak{p}}(z^1, z^2) = 0$ is the *minimal* local equation of Γ_ν at $\mathfrak{p}$. Put

$$h_p(D; \tau_p) = R_{\phi(p)}(D; \phi_p{}^1(\tau_p), \phi_p{}^2(\tau_p)).$$

Then $V_p(h_p(D))$ gives the intersection number of D and the branch Γ_p at $\mathfrak{p} = \phi(p)$. Putting $D \cdot \Gamma = \sum_p V_p(h_p(D)) \cdot p$, we introduce the *intersection divisor* $D \cdot \Gamma$ *of* D *and* Γ on $\mathbf{\Gamma}$. Obviously we have

$$I(D, \Gamma) = \deg (D \cdot \Gamma). \tag{5.1}$$

Now, inserting (4. 2) in (3. 34) and using $I(\Gamma, \Gamma) = \deg \mathfrak{d}$, we get immediately

$$\dim \{\Gamma\} = I(\Gamma, \Gamma) - \sum_{\nu=1}^{\mu} \pi_\nu + \mu - \sum_{\mathfrak{p}} l_{\mathfrak{p}} - q + i + j + 1.$$

Putting $\pi_\Gamma = \sum_{\nu=1}^{\mu} \pi_\nu - \mu + 1 + \frac{1}{2} \sum_p c_p$, we obtain therefore

$$\dim \{\Gamma\} = I(\Gamma, \Gamma) - \pi_\Gamma - q + \tfrac{1}{2} \sum_p c_p - \sum_{\mathfrak{p}} l_{\mathfrak{p}} + i + j + 2. \tag{5.2}$$

The quantity π_Γ is called the *virtual genus* [40] of Γ. Now we shall show that *the virtual genus* π_Γ *of* Γ *is determined uniquely by the homology class of* Γ.[41] For this purpose, we first introduce the notion of the *canonical cycles*.[42] Let

---

[40] Cf. Zariski [30], p. 30.

[41] This fact was suggested by A. Weil in his letter to the author.

[42] Weil [25].

$\{\mathfrak{U}_\lambda\}$ be a finite covering of $\mathfrak{M}$ by the "spherical" domains $\mathfrak{U}_\lambda$ each of which is covered by a single system of local coordinates $(z_\lambda^1, z_\lambda^2)$ such that each intersection $\mathfrak{U}_\lambda \cap \mathfrak{U}_\nu$ is simply connected (if it is not empty). Then, putting $dz_\nu^1 dz_\nu^2 = M_{\lambda\nu} dz_\lambda^1 dz_\lambda^2$, in $\mathfrak{U}_\lambda \cap \mathfrak{U}_\nu$, we get a system $\{M_{\lambda\nu}\}$ of the non-vanishing holomorphic functions $M_{\lambda\nu}$ defined respectively in $\mathfrak{U}_\lambda \cap \mathfrak{U}_\nu$ satisfying

$$M_{\lambda\nu} M_{\nu\lambda} = 1, \text{ in } \mathfrak{U}_\lambda \cap \mathfrak{U}_\nu; \qquad M_{\lambda\mu} M_{\mu\nu} M_{\nu\lambda} = 1, \text{ in } \mathfrak{U}_\lambda \cap \mathfrak{U}_\mu \cap \mathfrak{U}_\nu.$$

This system $\{M_{\lambda\nu}\}$ determines a *fibre bundle* $\mathfrak{F}$ over $\mathfrak{M}$ in the following manner: Consider the system $\{\mathfrak{U}_\lambda \times \mathfrak{E}\}$ of the direct products $\mathfrak{U}_\lambda \times \mathfrak{E}$ of $\mathfrak{U}_\lambda$ and the *projective line* $\mathfrak{E}$ and identify two points $\mathfrak{p} \times \zeta_\lambda \varepsilon \mathfrak{U}_\lambda \times \mathfrak{E}$ and $\mathfrak{p} \times \zeta_\nu \varepsilon \mathfrak{U}_\nu \times \mathfrak{E}$ if $\mathfrak{p} \varepsilon \mathfrak{U}_\lambda \cap \mathfrak{U}_\nu$ and $\zeta_\lambda / \zeta_\nu = M_{\lambda\nu}(\mathfrak{p})$. Then we get from $\{\mathfrak{U}_\lambda \times \mathfrak{E}\}$ a compact analytic manifold $\mathfrak{F}$ consisting of all points $\mathfrak{p} \times \zeta_\lambda$ which is obviously a fibre bundle over the base space $\mathfrak{M}$; the fibre of $\mathfrak{F}$ is $\mathfrak{E}$ and the transformation group in the fibre $\mathfrak{E}$ is the multiplicative group of complex numbers. Furthermore it can be easily verified that the analytic structure of $\mathfrak{F}$ relative to $\mathfrak{M}$ is determined uniquely by $\mathfrak{M}$. For an arbitrary point $P = \mathfrak{p} \times \zeta_\lambda$ of $\mathfrak{F}$, we denote by $\pi(P)$ the projection $\mathfrak{p}$ of P on the base space $\mathfrak{M}$. The subsets $\mathfrak{M}_0 = \{\mathfrak{p} \times 0 \mid \mathfrak{p} \varepsilon \mathfrak{M}\}$ and $\mathfrak{M}_\infty = \{\mathfrak{p} \times \infty \mid \mathfrak{p} \varepsilon \mathfrak{M}\}$ of $\mathfrak{F}$ are compact analytic sub-manifolds of complex dimension 2. Now, each $d \log M_{\lambda\nu}$ is a holomorphic simple differential in $\mathfrak{U}_\lambda \cap \mathfrak{U}_\nu$ and we have

$$d \log M_{\lambda\nu} + d \log M_{\nu\lambda} = 0, \quad \text{in } \mathfrak{U}_\lambda \cap \mathfrak{U}_\nu,$$

$$d \log M_{\lambda\mu} + d \log M_{\mu\nu} + d \log M_{\nu\lambda} = 0, \quad \text{in } \mathfrak{U}_\lambda \cap \mathfrak{U}_\mu \cap \mathfrak{U}_\nu.$$

Hence we can find a system $\{\Psi_\lambda\}$ of the 1-forms Ψ_λ of class C^∞ defined respectively in $\mathfrak{U}_\lambda$ such that[43]

$$\Psi_\lambda - \Psi_\nu = (1/2\pi i) d \log M_{\lambda\nu}, \quad \text{in } \mathfrak{U}_\lambda \cap \mathfrak{U}_\nu.$$

Putting $\mathrm{X} = d\Psi_\lambda$, we get a well defined 2-form X of class C^∞ on $\mathfrak{M}$, since $d\Psi_\lambda = d\Psi_\nu$ in $\mathfrak{U}_\lambda \cap \mathfrak{U}_\nu$. Considering $M_{\lambda\nu}(\mathfrak{p}) = M_{\lambda\nu}(\pi(P))$ as a function of $P = \mathfrak{p} \times \zeta_\lambda = \mathfrak{p} \times \zeta_\nu$ defined in $\pi^{-1}(\mathfrak{U}_\lambda \cap \mathfrak{U}_\nu)$, we have the identity $\mathrm{M}_{\lambda\nu} = \zeta_\lambda / \zeta_\nu$. Consequently we get $d \log M_{\lambda\nu} = d \log \zeta_\lambda - d \log \zeta_\nu$ and therefore

$$\Psi_\lambda - (1/2\pi i) d \log \zeta_\lambda = \Psi_\nu - (1/2\pi i) d \log \zeta_\nu, \quad \text{in } \pi^{-1}(\mathfrak{U}_\lambda \cap \mathfrak{U}_\nu),$$

where Ψ_λ, Ψ_ν are to be considered as 1-forms defined on $\pi^{-1}(\mathfrak{U}_\lambda)$, $\pi^{-1}(\mathfrak{U}_\nu)$, respectively. This result shows that $\Psi = \Psi_\lambda - (1/2\pi i) d \log \zeta_\lambda$ is a well defined 1-current on $\mathfrak{F}$. Now, we have $d \cdot d \log \zeta_\lambda = 2\pi i (\mathfrak{M}_0 - \mathfrak{M}_\infty)$, in

---

[43] Weil [24], p. 113.

$\pi^{-1}(\mathfrak{U}_\lambda)$. In fact, for an arbitrary 4-form Φ of class C^∞ with $|\Phi| \subset \pi^{-1}(\mathfrak{U}_\lambda)$, we get, putting $\zeta = \zeta_\lambda$ and $\eta = 1/\zeta_\lambda$,

$$\begin{aligned} d \cdot d \log \zeta[\Phi] = d \log \zeta[d\Phi] &= \int (d\zeta/\zeta) \cdot d\Phi \\ &= -\int d\{(d\zeta/\zeta) \cdot \Phi\} = -\lim_{\epsilon \to +0} \int_{\epsilon < |\zeta| < 1/\epsilon} d\{(d\zeta/\zeta) \cdot \Phi\} \\ &= \lim_{\epsilon \to +0} \{ \int_{|\zeta| = \epsilon} (d\zeta/\zeta) \cdot \Phi - \int_{|\eta| = \epsilon} (d\eta/\eta) \cdot \Phi \} \\ &= 2\pi i \{ \int_{\mathfrak{M}_0} \Phi - \int_{\mathfrak{M}_\infty} \Phi \} = 2\pi i (\mathfrak{M}_0 - \mathfrak{M}_\infty)[\Phi], \end{aligned}$$

proving the above formula. Hence we obtain

$$\text{X} = d\Psi + \mathfrak{M}_0 - \mathfrak{M}_\infty, \tag{5.3}$$

where X is to be considered as a 2-form on $\boldsymbol{\mathcal{F}}$. The formula (5.3) shows that *the cohomology class of the 2-form* X *on* $\mathfrak{M}$ *is uniquely determined by the analytic structure of* $\mathfrak{M}$. Moreover it can be shown that the cohomology class of X is the *second basic characteristic class*[44] of the manifold $\mathfrak{M}$. *By a canonical cycle on* $\mathfrak{M}$ *we shall mean a 2-cycle* $\mathfrak{k}$ *such that* $I(\mathfrak{k}, Z) = Z[\text{X}]$ *for all 2-cycles* Z *on* $\mathfrak{M}$.

Now we shall prove the formula

$$2\pi_\Gamma - 2 = I(\Gamma, \Gamma) + I(\mathfrak{k}, \Gamma). \tag{5.4}$$

For this purpose, we choose for each point $p \in \mathbf{\Gamma}$ a sufficiently small neighborhood U_p of p so that $\phi(U_p)$ is contained in one of the domains $\mathfrak{U}_\lambda$ which will be denoted by $\mathfrak{U}_{\lambda(p)}$: $\phi(U_p) \subset \mathfrak{U}_{\lambda(p)}$. Then, putting

$$m_{pq}(\tau) = M_{\lambda(p)\lambda(q)}(\phi(\tau)), \quad \text{in } U_p \cap U_q,$$

we get a system $\{m_{pq}\}$ of the holomorphic functions $\{m_{pq}\}$ satisfying the conditions

$$m_{pq} m_{qp} = 1, \quad \text{in } U_p \cap U_q; \qquad m_{pq} m_{qr} m_{rp} = 1, \quad \text{in } U_p \cap U_q \cap U_r.$$

Hence, by the same method as in the proof of Theorem 2.1, we can show that there exists a system $\{k_p\}$ of meromorphic functions k_p defined respectively in U_p satisfying $k_p/k_q = m_{pq}$, in $U_p \cap U_q$. Now, we associate with each point τ in $U_p \subset \mathbf{\Gamma}$ the point $P(\tau) = \phi(\tau) \times k_p(\tau)$ in $\mathfrak{U}_{\lambda(p)} \times \mathfrak{C}$. Then, since

$$k_p(\tau)/k_q(\tau) = m_{pq}(\tau) = M_{\lambda(p)\lambda(q)}(\phi(\tau)), \quad \text{in } U_p \cap U_q,$$

---

[44] See Chern [6], Theorem 8.

we have $\phi(\tau) \times k_p(\tau) = \phi(\tau) \times k_q(\tau)$, for $\tau \in U_p \cap U_q$. This shows that $\tau \to P(\tau) = \phi(\tau) \times k_p(\tau)$ is a well defined mapping which maps $\mathbf{\Gamma}$ onto a compact analytic curve $P(\mathbf{\Gamma})$ in $\mathfrak{J}$. Obviously we have $\pi(P(\tau)) = \phi(\tau)$ and therefore $\pi(P(\mathbf{\Gamma})) = \Gamma$. Hence we get, using (5. 3),

$$\Gamma[\mathrm{X}] = P(\mathbf{\Gamma})[\mathrm{X}] = ({*}\mathrm{X}, P(\Gamma)) = ({*}\mathfrak{M}_0 - {'}{*}\mathfrak{M}_\infty, P(\mathbf{\Gamma})),$$

yielding $I(\mathfrak{t}, \Gamma) = \Gamma[\mathrm{X}] = I(\mathfrak{M}_0 - \mathfrak{M}_\infty, P(\mathbf{\Gamma}))$. It can easily be verified that

$$I(\mathfrak{M}_0 - \mathfrak{M}_\infty, P(\mathbf{\Gamma})) = \sum_p V_p(k_p), \tag{5. 5}$$

where V_p denotes the valuation with the center p. In fact, a point $P(p)$ on $P(\mathbf{\Gamma})$ meets with $\mathfrak{M}_0$ if and only if $k_p(p) = 0$ (we write p instead of τ). The intersection number $I(\mathfrak{M}_0, P(\mathbf{\Gamma}))$ is therefore equal to the sum: $\sum_p I(\mathfrak{M}_0, P(U_p))$ extended over all p such that $k_p(p) = 0$, while we have

$$I(\mathfrak{M}_0, P(U_p)) = (1/2\pi i) \oint_p d \log k_p(\tau_p) = V_p(k_p).$$

Similarly, $P(p)$ meets with $\mathfrak{M}_\infty$ if and only if $1/k_p(p) = 0$ and, if $1/k_p(p) = 0$, we have

$$I(\mathfrak{M}_\infty, P(U_p)) = (1/2\pi i) \oint_p d \log(1/k_p(\tau_p)) = - V_p(k_p).$$

Hence we get (5. 5). Consequently we obtain

$$I(\mathfrak{t}, \Gamma) = \sum_p V_p(k_p). \tag{5. 6}$$

The differential σ_p depends on the choice of the local coordinates z^1, z^2 in $\mathfrak{U}_{\phi(p)}$. Using $z_{\lambda(p)}{}^1, z_{\lambda(p)}{}^2$ as the local coordinates in $\mathfrak{U}_{\phi(p)}$, we have by (3. 17)

$$k_q h_q \sigma_q = k_q [M_{\lambda(p)\lambda(q)}]_{z=\phi(\tau)} h_p \sigma_p = k_q m_{pq} h_p \sigma_p = k_p h_p \sigma_p, \quad \text{in } U_p \cap U_q.$$

This proves that $\xi = k_p h_p \sigma_p$ is a well defined differential on $\mathbf{\Gamma}$. The divisor (ξ) of ξ is represented as

$$(\xi) = \sum_p V_p(k_p) p + \mathfrak{d} - \mathfrak{c},$$

while we have

$$\deg(\xi) = \sum_{\nu=1}^{\mu} (2\pi_\nu - 2) = 2 (\sum_{\nu=1}^{\mu} \pi_\nu - \mu).$$

Hence we get

$$2\pi_\Gamma - 2 = \deg(\xi) + \deg \mathfrak{c} = \deg \mathfrak{d} + \sum_p V_p(k_p).$$

Combined with (5. 6), this yields immediately (5. 4). The formula (5. 4) shows clearly that π_Γ is determined uniquely by the homology class of Γ.

By a *canonical divisor* on $\mathfrak{M}$ we shall mean the divisor K of an arbitrary double differential $W = W_{12}(z^1, z^2)dz^1dz^2$ (defined in the whole of $\mathfrak{M}$) determined by means of the relation: $W_{12}(z^1, z^2) = R_{\mathfrak{p}}(K; z^1, z^2)$. Obviously any two canonical divisors K, K' are linearly equivalent to each other and therefore *the divisor class* $\{K\}$ *of a canonical divisor* K *is determined uniquely by* $\mathfrak{M}$, *if at least one double differential* $W \neq 0$ *ever exists on* $\mathfrak{M}$. Incidentally, the divisor of a double differential W will be denoted by (W). Now, assuming the existence of a double differential $W \neq 0$, we shall show that *any canonical divisor* $K = (W)$ *is a canonical cycle.* Let $\{\mathfrak{U}_\lambda\}$, $z_\lambda{}^\alpha$, $M_{\lambda\nu}$, etc. have the same meaning as above. Then, putting $W = R_\lambda(K; z_\lambda{}^1, z_\lambda{}^2)dz_\lambda{}^1dz_\lambda{}^2$, in $\mathfrak{U}_\lambda$, we have $M_{\lambda\nu} = R_\lambda(K)/R_\nu(K)$ in $\mathfrak{U}_\lambda \cap \mathfrak{U}_\nu$ and therefore

$$\Psi_\lambda - \Psi_\nu = (1/2\pi i)[d \log R_\lambda(K) - d \log R_\nu(K)], \quad \text{in } \mathfrak{U}_\lambda \cap \mathfrak{U}_\nu.$$

This shows that $\Phi = \Psi_\lambda - (1/2\pi i) d \log R_\lambda(K)$ is a well defined 1-form of class C^∞ in $\mathfrak{M} - |K|$. Now, let $Z = \sum c_jT_j$ be an arbitrary 2-cycle on $\mathfrak{M}$ represented as a linear combination of a finite number of differentiable simplices T_j such that each T_j is covered by a gingle neighborhood $\mathfrak{U}_{\lambda(j)}$ and that each $|dT_j|$ does not meet with $|K|$. Then we have

$$\begin{aligned} Z[\mathrm{X}] &= \sum c_jT_j[\mathrm{X}] = \sum c_jT_j[d\Psi_{\lambda(j)}] = -\sum c_jdT_j[\Psi_{\lambda(j)}] \\ &= -\sum c_jdT_j[\Phi] - (1/2\pi i)\sum c_jdT_j[d \log R_{\lambda(j)}(K)], \end{aligned}$$

while $\sum c_jdT_j = dZ = 0$ and, as (2.14) shows, $(1/2\pi i)dT_j[d \log R_{\lambda(j)}(K)] = -I(T_j, K)$. Hence we get $Z[\mathrm{X}] = I(K, Z)$, proving that K is a canonical cycle on $\mathfrak{M}$.

It follows from (5.4) that, by means of the canonical divisor K, the virtual genus π_Γ of Γ is represented as

$$2\pi_\Gamma - 2 = I(\Gamma, \Gamma) + I(K, \Gamma). \tag{5.7}$$

Now we shall prove the identity [45]

$$l_{\mathfrak{p}} = \tfrac{1}{2} \sum_{\phi(p)=\mathfrak{p}} c_p. \tag{5.8}$$

For that purpose, we first deduce from Theorem 3.2 the inequality

$$l_{\mathfrak{p}} \leqq \tfrac{1}{2} \sum_{\phi(p)=\mathfrak{p}} c_p. \tag{5.9}$$

Let $\mathfrak{p}$ be a fixed singular point of Γ. We denote by u a *vector* (or a *system*) $\{u_p \mid p \in \phi^{-1}(\mathfrak{p})\}$ consisting of polynomials $u_p = u_p(\tau_p)$ in τ_p of degree $\leqq c_p - 1$

---

[45] An arithmetic proof of this identity based on local considerations was given recently by D. Gorenstein. See Gorenstein [10].

and by $\mathfrak{I}_{\mathfrak{p}}$ the linear space consisting of all such vectors $u = \{u_p\}$. Obviously we have

$$\dim \mathfrak{I}_{\mathfrak{p}} = \sum_{\phi(p)=\mathfrak{p}} c_p.$$

Let (z^1, z^2) be a system of local coordinates with the origin $\mathfrak{p}$. Then every monomial $(z^1)^k(z^2)^l$ induces a holomorphic function $(\phi_p{}^1)^k(\phi_p{}^2)^l$ on each branch $\Gamma_p, p \varepsilon \phi^{-1}(\mathfrak{p})$. Denoting by $u_p{}^{(k,l)}$ the polynomial obtained from $(\phi_p{}^1)^k(\phi_p{}^2)^l$ by reducing it modulo $\tau_p{}^{c_p}$, we define the vectors $u^{(k,l)} = \{u_p{}^{(k,l)}\}$, $(k, l = 0, 1, 2, \cdots)$. Let $\mathfrak{L}_{\mathfrak{p}}$ be the subspace of $\mathfrak{I}_{\mathfrak{p}}$ generated by these induced vectors $u^{(k,l)}, k, l = 0, 1, 2, \cdots$. Then we have $\dim \mathfrak{L}_{\mathfrak{p}} = l_{\mathfrak{p}}$, since

$$\sigma_p{}^{(k,l)} = \sum_{\phi(p)=\mathfrak{p}} u_p{}^{(k,l)}\sigma_p + \text{holomorphic part}.$$

Now we introduce the bilinear form $b(u, v)$ of $u, v \,\varepsilon \mathfrak{I}_{\mathfrak{p}}$ defined as follows:

$$b(u, v) = \sum_{\phi(p)=\mathfrak{p}} \operatorname{Res}_p[u_p v_p \sigma_p].$$

Clearly $b(u, v)$ is symmetric and non-singular, while we have, by virtue of Theorem 3. 2,

$$b(u^{(k,l)}, v^{(h,j)}) = \sum_{\phi(p)=\mathfrak{p}} \operatorname{Res}_p[(\phi_p{}^1)^{k+h}(\phi_p{}^2)^{l+j}\sigma_p] = 0,$$

proving that $b(u, v) = 0$ for all $u, v \,\varepsilon \mathfrak{L}_{\mathfrak{p}}$. Hence we get

$$l_{\mathfrak{p}} = \dim \mathfrak{L}_{\mathfrak{p}} \leqq \tfrac{1}{2} \dim \mathfrak{I}_{\mathfrak{p}} = \tfrac{1}{2} \sum_{\phi(p)=\mathfrak{p}} c_p,$$

proving (5. 9).

By a suitable choice of the system of local coordinates (w, z) with the origin $\mathfrak{p}$ (we write here w, z instead of z^1, z^2), we may assume that $R_{\mathfrak{p}}(w, z)$ is a distinguished polynomial:

$$R_{\mathfrak{p}}(w, z) = w^m + B_1(z)w^{m-1} + \cdots + B_m(z)$$

and that the parametric representation of each branch $\Gamma_p, p \,\varepsilon\, \phi^{-1}(\mathfrak{p})$, is given by

$$w = \mathfrak{P}_p(\tau_p) = a_{p1}\tau_p + a_{p2}\tau_p + \cdots, \qquad z = \tau_p{}^{m_p}.$$

Each coefficient $B_\nu(z)$ in $R_{\mathfrak{p}}(w, v)$ is a holomorphic function:

$$B_\nu(z) = b_{\nu 1}z + b_{\nu 2}z^2 + \cdots.$$

Replacing these $B_\nu(z)$ by the polynomials

$$B^*{}_\nu(z) = b_{\nu 1}z + b_{\nu 2}z^2 + \cdots + b_{\nu N}z^N,$$

we construct the polynomial

$$R^*(w, z) = w^m + B^*_1(z) w^{m-1} + \cdots + B^*_m(z).$$

Then, as one readily infers, we can find for each $p \varepsilon \phi^{-1}(\mathfrak{p})$ and for an arbitrary given integer $M > 0$ a power series $\mathfrak{P}^*_p(\tau_p)$ such that

$$R^*(\mathfrak{P}^*_p(\tau_p), \tau_p^{m_p}) = 0, \qquad V_p(\mathfrak{P}^*_p(\tau_p) - \mathfrak{P}_p(\tau_p)) \geqq M,$$

if N is sufficiently large compared with M. Suppose M to be chosen sufficiently large. Then the functions

$$\mathfrak{P}^*_p(\zeta_p^k z^{1/m_p}), \qquad k = 0, 1, \cdots, m_p - 1, \quad p \varepsilon \phi^{-1}(\mathfrak{p})$$

are all different to each other, where $\zeta_p = \exp(2\pi i/m_p)$, and we have

$$R^*(w, z) = \prod_p \prod_{k=0}^{m_p-1} [w - \mathfrak{B}^*_p(\zeta_p^k z^{1/m_p})].$$

Hence, for each $p \varepsilon \phi^{-1}(\mathfrak{p})$, the expression $w = \mathfrak{B}^*_p(\tau_p)$, $z = \tau_p^{m_p}$ represents an irreducible branch Γ^*_p and the curve Γ^* defined by the equation $R^*(w, z) = 0$ consists in $\mathfrak{U}_\mathfrak{p}$ of the irreducible branches $\Gamma^*_p, p \varepsilon \phi^{-1}(\mathfrak{p})$. Denote by $l^*_\mathfrak{p}$, c^*_p the integers corresponding to $l_\mathfrak{p}$, c_p associated with Γ^*, Γ^*_p. c^*_p is the order of the pole $\tau_p = 0$ of the differential

$$\sigma^*_p = [m_p \tau_p^{m_p - 1} / \partial_1 R^*(\mathfrak{P}^*_p(\tau_p), \tau_p^{m_p})] d\tau_p$$

and $l^*_\mathfrak{p}$ is the number of formal sums $\sum_{\phi(p)=\mathfrak{p}} (\mathfrak{P}^*_p)^k \tau_p^{l m_p} \sigma^*_p$, $(k, l = 0, 1, 2, \cdots)$ which are independent modulo 0. Hence, if M and N are chosen sufficiently large, we have $c^*_p = c_p$, $l^*_\mathfrak{p} = l_\mathfrak{p}$. Supposing (w, z) as the inhomogeneous coordinates on the projective plane, we can consider Γ^* as a plane algebraic curve defined by the algebraic equation $R^*(w, z) = 0$. In order to prove (5.8), it is sufficient therefore to show

$$(5.8)^* \qquad l^*_\mathfrak{p} = \tfrac{1}{2} \sum_{\phi(p)=\mathfrak{p}} c^*_p$$

with respect to this plane algebraic curve Γ^*. Now we apply the formula (5.2) to Γ^*. Then we get

$$\dim\{\Gamma^*\} = I(\Gamma^*, \Gamma^*) - \pi_{\Gamma^*} + \sum_\mathfrak{p} [\tfrac{1}{2} \sum_{\phi(p)=\mathfrak{p}} c^*_p - l^*_\mathfrak{p}] + i + 2,$$

since, in this case, $q = j = 0$. On the other hand, denoting by n the degree of Γ^*, we have

$$\dim\{\Gamma^*\} = \tfrac{1}{2}(n^2 + 3n + 2), \qquad \pi_{\Gamma^*} = \tfrac{1}{2}(n^2 - 3n + 2), \qquad I(\Gamma^*, \Gamma^*) = n^2.$$

Hence we get $\sum_\mathfrak{p} [\frac{1}{2} \sum_{\phi(p)=\mathfrak{p}} c^*_p - l^*_\mathfrak{p}] + i = 0$. Combined with the inequality $i \geqq 0$ and (5.9), this yields immediately $(5.8)^*$, q. e. d.

The space $\Xi(\Gamma, -\mathfrak{d})$ was defined as the set of all differentials ξ on $\mathbf{\Gamma}$ satisfying

$$\xi \equiv \sum_{\mathfrak{p}} \sum_{k,l} a_{\mathfrak{p}kl} \sum_{\phi(p)=\mathfrak{p}} (\phi_p{}^1)^k (\phi_p{}^2)^l h_p \sigma_p \ (\mathfrak{d}). \tag{5.10}$$

Clearly this condition implies

$$(\xi) - \mathfrak{d} + \mathfrak{c} \geqq 0. \tag{γ}$$

Now, let ξ be an arbitrary differential satisfying (γ). Then, for each singular point $\mathfrak{p}$ of Γ, the formal sum $\sum_{\phi(p)=\mathfrak{p}} h_p{}^{-1}\xi$ can be written as

$$\sum_{\phi(p)=\mathfrak{p}} h_p{}^{-1}\xi = \sum_{\phi(p)=\mathfrak{p}} u_p\sigma_p + \text{holomorphic part}, \tag{5.11}$$

where $u = \{u_p\} \varepsilon\, \mathfrak{l}_{\mathfrak{p}}$. Obviously this ξ satisfies (5.10) if and only if $u = \{u_p\}$ belongs to $\mathfrak{L}_{\mathfrak{p}}$ for each singular point $\mathfrak{p}$ of Γ. The identity (5.8) shows that $\dim \mathfrak{L}_{\mathfrak{p}} = \frac{1}{2} \dim \mathfrak{l}_{\mathfrak{p}}$; consequently, an arbitrary element $u \,\varepsilon\, \mathfrak{l}_{\mathfrak{p}}$ belongs to $\mathfrak{L}_{\mathfrak{p}}$ if and only if u satifies $b(u, v) = 0$ for all $v \,\varepsilon\, \mathfrak{L}_{\mathfrak{p}}$, while, for $u = \{u_p\}$ appearing in (5.11), we have

$$b(u, v) = \sum_{\phi(p)=\mathfrak{p}} \operatorname{Res}_p[v_p h_p{}^{-1}\xi].$$

Hence ξ satisfies (5.10) if and only if

$$\sum_{\phi(p)=\mathfrak{p}} \operatorname{Res}_p[u_p{}^{(k,l)} h_p{}^{-1}\xi] = 0, \qquad (k, l = 0, 1, 2, \cdots)$$

for each singular point $\mathfrak{p}$ of Γ. Thus we see that *the space* $\Xi(\Gamma, -\mathfrak{d})$ *consists of all differentials* ξ *on* $\mathbf{\Gamma}$ *satisfying*

$$(\xi) - \mathfrak{d} + \mathfrak{c} \geqq 0 \tag{γ}$$

and

$$\sum_{\phi(p)=\mathfrak{p}} \operatorname{Res}_p[(\phi_p{}^1)^k (\phi_p{}^2)^l h_p{}^{-1}\xi] = 0, \qquad (k, l = 0, 1, 2, \cdots) \tag{δ}$$

for each singular point $\mathfrak{p}$ *of* Γ.

Now, using the identity (5.8), we get from (5.2) immediately the following

THEOREM 5.1 (THE THEOREM OF RIEMANN-ROCH). *The dimension of the divisor class* $\{\Gamma\}$ *of a (reducible or irreducible) curve* Γ *is given by*

$$\dim\{\Gamma\} = I(\Gamma, \Gamma) - \pi_\Gamma - q + i + j + 2, \tag{5.12}$$

where, q *is the irregularity of* $\mathfrak{M}$, π_Γ *is the virtual genus of* Γ, i *is the dimension of the linear space* $\Xi(\Gamma, -\mathfrak{d})$ *and* j *is the number of linearly independent simple differentials* A *of the first kind on* $\mathfrak{M}$ *satisfying* $Q_f[\Omega \cdot A] = 0$ *for all* $f \,\varepsilon\, \mathfrak{f}(\Gamma, \mathfrak{d})$.

The number g of linearly independent double differentials of the first kind on $\mathfrak{M}$ is called the *geometric genus* of $\mathfrak{M}$. For an arbitrary divisor D on $\mathfrak{M}$, we denote by $\mathfrak{W}(D)$ the linear space consisting of all double differentials W such that $(W) + D \geqq 0$. Then we have

THEOREM 5. 2 (RIEMANN-ROCH'S INEQUALITY). *We have*

$$(5.13) \qquad \dim\{\Gamma\} + \dim \mathfrak{W}(-\Gamma) \geqq I(\Gamma, \Gamma) - \pi_\Gamma - q + g + 2.$$

Proof. For every double differential $B = B_{12}(z)\,dz^1dz^2$ of the first kind, $\xi = B_{12}(\phi_p)h_p\sigma_p$ is a well defined differential on $\mathbf{\Gamma}$, as one readily infers by (3. 17); furthermore, this differential satisfies clearly the above conditions (γ) and (δ). Thus $B \rightarrow \xi = B_{12}(\phi_p)h_p\sigma_p$ is a linear mapping which maps the space $\mathfrak{W}(0)$ of all double differentials of the first kind *into* the space $\Xi(\Gamma, -\mathfrak{d})$. Obviously $\xi = B_{12}(\phi_p)h_p\sigma_p$ vanishes identically on $\mathbf{\Gamma}$ if and only if $B \varepsilon \mathfrak{W}(-\Gamma)$. Consequently we get

$$i = \dim \Xi(\Gamma, -\mathfrak{d}) \geqq g - \dim \mathfrak{M}(-\Gamma).$$

Combined with (5. 12), this yields immediately the inequality (5. 13), q. e. d.

In what follows we assume that *there exists on $\mathfrak{M}$ at least one double differential W_0 not vanishing identically and put $K = (W_0)$*. Then, for an arbitrary fixed divisor D, every double differential $W \varepsilon \mathfrak{W}(D)$ corresponds one-to-one to a meromorphic function $F \varepsilon \mathfrak{F}(K + D)$ by means of the relation: $W = FW_0$; consequently we have

$$(5.14) \qquad \dim \mathfrak{W}(D) = \dim\{K + D\}.$$

Using this identity, we get from Theorem 5. 2 immediately the following

THEOREM 5. 3 (RIEMANN-ROCH'S INEQUALITY). *We have*

$$(5.15) \qquad \dim\{\Gamma\} + \dim\{K - \Gamma\} \geqq I(\Gamma, \Gamma) - \pi_\Gamma + g - q + 2.$$

Now, we introduce the notion of the *deficiency* of an arbitrary divisor D on the (reducible or irreducible) curve Γ. Assume first that D and Γ have no component in common and denote by $\mathfrak{f}(\Gamma, D)$ the space consisting of all meromorphic functions f on $\mathbf{\Gamma}$ satisfying

$$(\alpha)_D \qquad f \equiv 0(-D \cdot \Gamma)$$

and

$$(\beta)_D \qquad \sum_{\phi(p)=\mathfrak{p}} \mathrm{Res}_p[(\phi_p{}^1)^k(\phi_p{}^2)^l f h_p^{\cdot}(D)\sigma_p] = 0, \qquad (k, l = 0, 1, 2, \cdots)$$

for each singular point $\mathfrak{p}$ of Γ. Then every meromorphic function $F \varepsilon \mathfrak{F}(D)$

induces on Γ the meromorphic function $f(\tau) = F(\phi_p(\tau))$ belonging to $\mathfrak{f}(\Gamma, D)$ and the induced function $f(\tau)$ vanishes identically if and only if F belongs to $\mathfrak{F}(D - \Gamma)$. Hence we have

$$\dim \mathfrak{f}(\Gamma, D) \geqq \dim \{D\} - \dim \{D - \Gamma\}.$$

Now we define the *deficiency*[46] $\operatorname{def}(D/\Gamma)$ of the divisor D on the curve Γ as

$$\operatorname{def}(D/\Gamma) = \dim \mathfrak{f}(\Gamma, D) - [\dim \{D\} - \dim \{D - \Gamma\}].$$

As one readily infers, if *D' is linearly equivalent* to D, then $\operatorname{def}(D'/\Gamma) = \operatorname{def}(D/\Gamma)$. In case D and Γ have some components in common, we define $\operatorname{def}(D/\Gamma)$ as $\operatorname{def}(D/\Gamma) = \operatorname{def}(D'/\Gamma)$ by means of a divisor $D' \approx D$ having no component in common with Γ, assuming the existence of such a divisor D' (this is the case if $\mathfrak{M}$ is an algebraic surface). It is to be noted here that $\operatorname{def}(\Gamma/\Gamma)$ *coincides with the characteristic deficiency* $chd(\Gamma)$ *of* Γ, whenever $\operatorname{def}(\Gamma/\Gamma)$ can be defined. In fact, if there exists a divisor $D \approx \Gamma$ such that D and Γ have no component in common, then we have a meromorphic function F with the divisor $(F) = D - \Gamma$, and, putting

$$R_{\mathfrak{p}}(D) = R_{\mathfrak{p}}F, \tag{5.16}$$

we get a meromorphic function $R_{\mathfrak{p}}(D)$ representing D in $\mathfrak{U}_{\mathfrak{p}}$. It follows from (5.16) that the functions $h_p(D) = R_{\phi(p)}(D; \phi_p)$ satisfy

$$h_p(D)/h_q(D) = R_{\phi(p)}/R_{\phi(q)}, \quad \text{in } U_p \cap U_q,$$

proving that $\{h_p(D)\}$ can be used as a system $\{h_p\}$ defining the characteristic divisor $\mathfrak{d}$. Thus we see that $D \cdot \Gamma$ is the characteristic divisor $\mathfrak{d}$ so that $\mathfrak{f}(\Gamma, D)$ coincides with $\mathfrak{f}(\Gamma, \mathfrak{d})$. Hence we get

$$\operatorname{def}(D/\Gamma) = \dim \mathfrak{f}(\Gamma, \mathfrak{d}) - \dim \{\Gamma\} + 1,$$

proving that $\operatorname{def}(\Gamma/\Gamma) = chd(\Gamma)$.

The difference

$$\sup \{\Gamma\} = \dim \{\Gamma\} + \dim \{K - \Gamma\} - I(\Gamma, \Gamma) + \pi_\Gamma + q - g - 2$$

is called the *superabundance* of the divisor class $\{\Gamma\}$. From our deduction of Riemann-Roch's inequality it follows immediately that

$$\sup \{\Gamma\} = i - [g - \dim \{K - \Gamma\}] + q - chd(\Gamma).$$

Now, assuming the existence of a canonical divisor K such that K and Γ have no component in common, we shall show that $i-[g-\dim\{K-\Gamma\}]$ is equal to $\operatorname{def}(K/\Gamma)$. Let $W=W_{12}dz^1dz^2$ be the double differential with $(W)=K$ and consider the one-to-one correspondence $f\rightarrow\xi=fW_{12}(\phi_p)h_p\sigma_p$ between meromorphic functions f and differentials ξ on $\boldsymbol{\Gamma}$. Since the divisor $(W_{12}(\phi_p)h_p\sigma_p)$ of $W_{12}h_p\sigma_p$ is $K\cdot\Gamma+\mathfrak{d}-\mathfrak{c}$, ξ satisfies $(\xi)-\mathfrak{d}+\mathfrak{c}\geqq 0$ if and only if $(f)+K\cdot\Gamma\geqq 0$. Again, since $W_{12}(\phi_p)=h_p(K)$, ξ satisfies the condition

---

[46] Cf. Goldman [9], § 6. In the terminology of Italian geometers, $\operatorname{def}(D/\Gamma)$ is the deficiency of the linear series on Γ cut out by the complete linear system $|D|$.

$$\sum_{\phi(p)=\mathfrak{p}}\operatorname{Res}_p[(\phi_p{}^1)^k(\phi_p{}^2)^l h_p{}^{-1}\xi]=0$$

if and only if f satisfies

$$\sum_{\phi(p)=\mathfrak{p}}\operatorname{Res}_p[(\phi_p{}^1)^k(\phi_p{}^2)^l f h_p(K)\sigma_p]=0.$$

Thus ξ belongs to $\Xi(\Gamma,-\mathfrak{d})$ if and only if $f\,\varepsilon\,\mathfrak{f}(\Gamma,K)$. Hence we have $i=\dim\Xi(\Gamma,-\mathfrak{d})=\dim\mathfrak{f}(\Gamma,K)$, while, as (5.14) shows, $g=\dim\mathfrak{W}(0)=\dim\{K\}$. We get therefore

$$i-[g-\dim\{K-\Gamma\}]=\dim\mathfrak{f}(\Gamma,K)-[\dim\{K\}-\dim\ \{K-\Gamma\}],$$

proving that $i-[g-\dim\{K-\Gamma\}]$ is equal to $\operatorname{def}(K/\Gamma)$. Thus we obtain the following

THEOREM 5.4 (G. Castelnuovo[47]). *The superabundance of the divisor class of an arbitrary curve* Γ *is represented as*

$$\sup\{\Gamma\}=\operatorname{def}(K/\Gamma)+q-chd(\Gamma). \tag{5.17}$$

6. Double differentials. In this Section we shall determine the dimension of the linear space $\mathfrak{W}(\Gamma)$ consisting of all double differentials on $\mathfrak{M}$ which are multiples of $-\Gamma$ in the sense that $(W)+\Gamma\geqq 0$. As was shown in Section **3**, the residue $\xi=\mathfrak{R}(W)=R_pW_{12}(\phi_p)\sigma_p$ of an arbitrary double differential $W\,\varepsilon\,\mathfrak{W}(\Gamma)$ is a well defined differential on $\boldsymbol{\Gamma}$. This differential $\xi=\mathfrak{R}(W)$ satisfies obviously

$$(\xi)+\mathfrak{c}\geqq 0; \tag*{$(\gamma)_0$}$$

again, as one readily infers by Theorem 3.2, $\xi=\mathfrak{R}(W)$ satisfies

$$\sum_{\phi(p)=\mathfrak{p}}\operatorname{Res}_p[(\phi_p{}^1)^k(\phi_p{}^2)^l\xi]=0, \qquad (k,l=0,1,2,\cdots) \tag*{$(\delta)_0$}$$

for each singular $\mathfrak{p}$ of Γ. *We denote by* $\Xi(\Gamma)$ *the space consisting of all differentials* ξ *on* $\boldsymbol{\Gamma}$ *satisfying* $(\gamma)_0$ *and* $(\delta)_0$. The mapping $W\rightarrow\xi=\mathfrak{R}(W)$

---

[47] This theorem was proved by G. Castelnuovo in the case that $\mathfrak{M}$ is an algebraic surface. Cf. Castelnuovo [4]; see also Zariski [30], p. 71.

is a linear transformation which maps $\mathfrak{M}(\Gamma)$ *into* $\Xi(\Gamma)$. Clearly $\xi = \mathfrak{R}(W)$ vanishes identically if and only if W is a differential of the first kind.

Now, suppose an arbitrary differential $\xi \varepsilon \Xi(\Gamma)$ as given. Then, does there exist a double differential $W \varepsilon \mathfrak{W}(\Gamma)$ such that $\mathfrak{R}(W) = \xi$? In order to solve this question, we associate with ξ the 3-current T_ξ defined as

$$T_\xi[X] = \int_\Gamma \xi \cdot \sum_{\alpha} X_{\alpha *} d\bar{\phi}_p{}^{\alpha}, \tag{6.1}$$

where X is a "variable" 1-form of class C^∞ and the value of the integral in the right-hand side is to be interpreted as the Cauchy principal value, in case it does not converge absolutely (see (3.15)). T_ξ is obviously a pure 3-current of type 1. Again we have

$$dT_\xi = 0. \tag{6.2}$$

In fact, for an arbitrary function $\Psi(z)$ of class C^∞, we have

$$dT_\xi[\Psi] = T_\xi[d\Psi] = \int_\Gamma \xi \cdot d\Psi(\phi_p) = -\int_\Gamma \mathrm{d}(\Psi(\phi_p)\xi),$$

while $\Psi(\phi_p)\xi$ is a 1-form of class C^∞ in $\Gamma - |\mathfrak{c}|$. Hence we get

$$dT_\xi[\Psi] = \sum_{\mathfrak{p}} \sum_{\phi(p)=\mathfrak{p}} \lim_{\epsilon \to +0} \int_{|\tau_p|=\epsilon} \Psi(\phi_p)\xi,$$

where $\sum_{\mathfrak{p}}$ denotes the sum extended over all singular points $\mathfrak{p}$ of Γ; while, using Lemma 3.1, we get from $(\delta)_0$ the identity

$$\sum_{\phi(p)=\mathfrak{p}} \lim_{\epsilon \to +0} \int_{|\tau_p|=\epsilon} \Psi(\phi_p)\xi = 0.$$

Hence, we obtain $dT_\xi[\Psi] = 0$, proving (6.2). Now, if there exists $W \varepsilon \mathfrak{W}(\Gamma)$ such that $\mathfrak{R}(W) = \xi$, then as (3.15) shows, we have

$$dW = 2\pi i T_\xi \tag{6.3}$$

and therefore

$$HT_\xi = 0. \tag{6.4}$$

Now, assuming conversely that T_ξ satisfies (6.4), we shall prove the existence of $W \varepsilon \mathfrak{W}(\Gamma)$ with $\mathfrak{R}(W) = \xi$. For this purpose, we introduce the 2-current θ_ξ defined as $\theta_\xi = (d\Lambda G + i\delta G)T_\xi$. By virtue of Theorem 1.4, θ_ξ is a holomorphic 2-form in $\mathfrak{M} - \Gamma$. In order to show that θ_ξ is meromorphic on Γ, we first consider an arbitrary simple point $\mathfrak{p}$ of Γ. Choose the system of local coordinates (w, z) in a neighborhood $\mathfrak{U}_{\mathfrak{p}}$ of $\mathfrak{p}$ so that $R_{\mathfrak{p}} = w$ and that $z = \tau_p$ on Γ and put $\xi = \xi_1(\tau_p)d\tau_p$ in U_p, where $p = \phi^{-1}(\mathfrak{p})$.

By virtue of $(\gamma)_0$, $\xi_1(\tau_p)$ is a holomorphic function in U_p. Hence, putting $W_{\mathfrak{p}}(w, z) = (dw/w) \cdot \xi_1(z)dz$, we get a double differential $W_{\mathfrak{p}}(w, z)$ defined in $\mathfrak{U}_{\mathfrak{p}}$. Considering $W_{\mathfrak{p}}$ as a 2-current in $\mathfrak{U}_{\mathfrak{p}}$, we get by (3. 10), $dW_{\mathfrak{p}} = 2\pi i T_\xi$, in $\mathfrak{U}_{\mathfrak{p}}$, since, in this case, $\sigma_p = d\tau_p$. On the other hand, θ_ξ is a pure 2-current of type 0 and satisfies

$$d\theta_\xi = iT_\xi, \tag{6. 5}$$

as Theorem 1. 4 shows. Hence we get $d(2\pi\theta_\xi - W_{\mathfrak{p}}) = 0$, in $\mathfrak{U}_{\mathfrak{p}}$. By virtue of Theorem 1. 3, we infer therefore that $B_{\mathfrak{p}} = 2\pi\theta_\xi - W_{\mathfrak{p}}$ is a holomorphic 2-form in $\mathfrak{U}_{\mathfrak{p}}$. Consequently $2\pi\theta_\xi$ is a meromorphic 2-form in $\mathfrak{U}_{\mathfrak{p}}$; moreover, since $\sigma_p = d\tau_p = dz$ and $2\pi R_{\mathfrak{p}}\theta_\xi = \xi_1(z)dwdz + R_{\mathfrak{p}}B_{\mathfrak{p}}$, we get

$$2\pi R_{\mathfrak{p}}\theta_{\xi 12}(\phi_p)\sigma_p = \xi, \qquad (p = \phi^{-1}(\mathfrak{p})). \tag{6. 6}$$

Now, assume $\mathfrak{p}$ to be a singular point of Γ. Then it follows from the above result that $R_{\mathfrak{p}}\theta_\xi$ is holomorphic in $\mathfrak{U}_{\mathfrak{p}} - \mathfrak{p}$. Hence, by virtue of a theorem of Hartogs,[48] $R_{\mathfrak{p}}\theta_\xi$ must be holomorphic in the whole of $\mathfrak{U}_{\mathfrak{p}}$. Thus $R_{\mathfrak{p}}\theta_\xi$ is holomorphic in $\mathfrak{U}_{\mathfrak{p}}$ for every $\mathfrak{p} \varepsilon \mathfrak{M}$, proving that θ_ξ is a double differential belonging to $\mathfrak{W}(\Gamma)$. Moreover it follows from (6. 6) that $2\pi\theta_\xi$ satisfies $\mathfrak{R}(2\pi\theta_\xi) = \xi$. Thus the existence of $W \varepsilon \mathfrak{W}(\Gamma)$ satisfying $\mathfrak{R}(W) = \xi$ is proved.

For an arbitrary simple differential $A = \sum A_\alpha(z)dz^\alpha$ of the first kind, we put

$$A_\Gamma = \sum_\alpha A_\alpha(\phi_p)d\phi_p{}^\alpha.$$

A_Γ is the *differential induced on* $\boldsymbol{\Gamma}$ *by* A. It is clear that HT_ξ vanishes if and only if $T_\xi[\Phi] = 0$ for all harmonic 1-form Φ of the first kind, while every harmonic 1-form Φ of the first kind can be written as $\Phi = A' + \bar{A}$ by means of two simple differentials A', A of the first kind. Hence HT_ξ vanishes if and only if

$$T_\xi[\bar{A}] = \int_\Gamma \xi \cdot \bar{A}_\Gamma = 0$$

for all simple differentials A of the first kind. Thus we obtain the following

THEOREM 6. 1. *The linear mapping* $W \rightarrow \xi = \mathfrak{R}(W)$ *maps the space* $\mathfrak{W}(\Gamma)$ *onto the subspace of* $\Xi(\Gamma)$ *consisting of all differentials* $\xi \varepsilon \Xi(\Gamma)$ *satisfying*

$$\int_\Gamma \xi \cdot \bar{A}_\Gamma = 0 \tag{6. 7}$$

---

[48] Hartogs [11].

for all simple differentials A of the first kind on $\mathfrak{M}$. Every $W \varepsilon \mathfrak{W}(\Gamma)$ can be represented by means of $\xi = \mathfrak{R}(W)$ as

$$W = 2\pi\theta_\xi + B, \tag{6.8}$$

where B is a double differential of the first kind. Conversely, for every $\xi \varepsilon \Xi(\Gamma)$ satisfying (6.7), the double differential W given by (6.8) belongs to $\mathfrak{W}(\Gamma)$ and satisfies $\mathfrak{R}(W) = \xi$.

Denote by k the number of linearly independent simple differentials A of the first kind such that A_Γ vanishes identically. Then the number of linearly independent *induced* differentials A_Γ on $\mathbf{\Gamma}$ is equal to $q - k$. Moreover each A_Γ is a differential of the first kind and therefore is contained in $\Xi(\Gamma)$. Hence the number of linearly independent differentials $\xi \varepsilon \Xi(\Gamma)$ satisfying (6.7) is equal to $\dim \Xi(\Gamma) - q + k$. By virtue of Theorem 6.1, we infer therefore that the dimension of the space $\mathfrak{W}(\Gamma)$ is given by

$$\dim \mathfrak{W}(\Gamma) = g + \dim \Xi(\Gamma) - q + k. \tag{6.9}$$

The dimension of $\Xi(\Gamma)$ is closely connected with the virtual genus π_Γ of Γ. Consider an arbitrary system $\{\xi_p \,|\, p \varepsilon |\mathfrak{c}|\}$ of differentials ξ_p defined respectively at $p \varepsilon |\mathfrak{c}|$ and denote by $\sum_p \xi_p$ the formal sum of ξ_p introduced in Section **4**. By virtue of (5.8), the number of $\sum_p \xi_p$ satisfying $(\gamma)_0$ and $(\delta)_0$ which are independent modulo 0 is given by $\frac{1}{2} \sum_p c_p$, while, for arbitrary $\sum_p \xi_p$, there exists a differential ξ on $\mathbf{\Gamma} = \sum_{\nu=1}^{\mu} \mathbf{\Gamma}_\nu$ satisfying $\xi \equiv \sum_p \xi_p \ (0)$ if and only if

$$\sum_p^{(\nu)} \operatorname{Res}_p[\xi_p] = 0, \qquad (\nu = 1, 2, \cdots, \mu), \tag{6.10}$$

where $\sum_p^{(\nu)}$ denotes the sum extended over all $p \varepsilon |\mathfrak{c}| \cap \mathbf{\Gamma}_\nu$. On the other hand, for arbitrarily preassigned values ρ_p, $p \varepsilon |\mathfrak{c}|$, satisfying $\sum_{\phi(p)=\mathfrak{p}} \rho_p = 0$ for each singular point $\mathfrak{p}$ of Γ, we can choose $\sum_p \xi_p$ satisfying $(\gamma)_0$ and $(\delta)_0$ so that $\operatorname{Res}_p[\xi_p] = \rho_p$ for each $p \varepsilon |\mathfrak{c}|$. Decompose Γ into the sum $\Gamma = \sum_{s=1}^{m} \Gamma^{(s)}$ of *connected components* $\Gamma^{(s)} = \sum_{\nu=\mu(s-1)+1}^{\mu(s)} \Gamma_\nu$, where $\mu(0) = 0$ and $\mu(m) = \mu$. Then it fofllows from the above results that one can choose $\sum_p \xi_p$ satisfying

$(\gamma)_0$ and $(\delta)_0$ so that each $\overset{(\nu)}{\sum_{p}} \mathrm{Res}_p[\xi_p]$ $(\mu(s-1) < \nu < \mu(s),\ s=1,2,\cdots,m)$ takes an arbitrarily preassigned value, while $(\gamma)_0$ and $(\delta)_0$ imply

$$\overset{(\mu(s))}{\sum_{p}} \mathrm{Res}_p[\xi_p] = -\sum_{\mu(s-1)<\nu<\mu(s)} \overset{(\nu)}{\sum_{p}} \mathrm{Res}_p[\xi_p].$$

Hence the number of linearly independent conditions involved in (6.10) is $\mu - m$, while the number of linearly independent differentials satisfying $\xi \equiv 0(0)$ is $\sum_{\nu=1}^{\mu} \pi_\nu$. Consequently we get

$$\dim \Xi(\Gamma) = \sum_{\nu=1}^{\mu} \pi_\nu + \tfrac{1}{2}\sum c_p - \mu + m = \pi_\Gamma + m - 1.$$

Inserting this into (6.9), we obtain the following

THEOREM 6.2. *The dimension of the space* $\mathfrak{W}(\Gamma)$ *is given by*

$$\dim \mathfrak{W}(\Gamma) = g + \pi_\Gamma + m - 1 - q + k, \tag{6.11}$$

where π_Γ *is the virtual genus of* Γ, m *is the number of connected component of* Γ *and* k *is the number of linearly independent simple differentials* A *of the first kind such that* A_Γ *vanishes identically on* $\boldsymbol{\Gamma}$.

It is obvious that A_Γ vanishes identically if and only if the simple Picard integral $\int^z A$ is constant on each connected component of Γ. Hence *k is equal to the number of linearly independent simple Picard integrals of the first kind which are constant on each connected component of* Γ. The difference $dd(\Gamma) = g + \dim \Xi(\Gamma) - \dim \mathfrak{W}(\Gamma)$ is called the differential deficiency[49] of Γ. The formula (6.9) shows that

$$dd(\Gamma) = q - k; \tag{6.12}$$

thus $dd(\Gamma)$ *is always not greater than* q. Let Δ be an arbitrary curve which is *homologous* to Γ. *Then* A_Δ *is identically zero if and only if* A_Γ *is identically zero,* A being an arbitrary simple differential of the first kind. In fact, denoting by $\tau = u + iv$ the local uniformization variable on $\boldsymbol{\Gamma}$, we have

$$\int_\Gamma A \cdot \bar{A} = \int_{\boldsymbol{\Gamma}} A_\Gamma \cdot \bar{A}_\Gamma = -2i \int_{\boldsymbol{\Gamma}} |\, A_\Gamma / d\tau\, |^2 du dv,$$

---

[49] Goldman [9], § 6. In the terminology of Italian geometers, $dd(\Gamma)$ is the deficiency of the linear series on Γ cut out by the adjoint system of the complete linear system $|\Gamma|$. See the formula (6.17) below.

proving that A vanishes identically if and only if $\int_\Gamma A \cdot \bar{A}$ is zero, while, since $d[A \cdot \bar{A}] = 0$, we have

$$\int_\Delta A \cdot \bar{A} = \int_\Gamma A \cdot \bar{A}.$$

Hence A_Δ is zero if and only if A_Γ is zero. This shows that the number k of indepenedent simple differentials A of the first kind satisfying $A_\Gamma = 0$ is determined uniquely by the homology class of Γ so that the *differential deficiency* $dd(\Gamma)$ *of* Γ *is uniquely determined by the homology class of* Γ. Now, it follows from (6.11) that

$$m = \dim \mathfrak{W}(\Gamma) - \pi_\Gamma - g + q + 1 - k,$$

proving that the *number* m *of the connected component of* Γ *is determined uniquely by the divisor class* $\{\Gamma\}$ *of* Γ.[50]

THEOREM 6.3. *If* Γ *is connected and* $\dim\{\Gamma\} \geqq 2$, *then the differential deficiency* $dd(\Gamma)$ *of* Γ *is equal to the irregularity* q *of* $\mathfrak{M}$: $dd(\Gamma) = q$.

Proof. By virtue of (6.12), it is sufficient to show that any simple Picard integral $\int^z A$ of the first kind is not a constant on Γ unless A vanishes identically. By hypothesis, there exists on $\mathfrak{M}$ at least one non-constant meromorphic function F such that $(F) + \Gamma \geqq 0$. Putting $(F - t) = \Gamma_t - \Gamma$, t being a complex parameter, we get a linear pencil $\{\Gamma_t\}$. Γ_t is a curve (i.e. a positive divisor without multiple component) unless t takes one of a finite number of special values $t_1, t_2, \cdots, t_s$; moreover, by the above result, Γ_t is connected for $t \neq t_\nu$. Now, if $\int^z A$ is constant on Γ, then $\int^z A$ is constant on each Γ_t, $t \neq t_\nu$, since Γ_t is homologous to Γ. Let $\alpha(t)$ be the value of $\int^z A$ on Γ_t, $t \neq t_\nu$. Then $\alpha(t)$ is an additive holomorphic function of t for $t \neq t_\nu$ $(\nu = 1, 2, \cdots, s)$ ($\alpha(t)$ is holomorphic also for $t = \infty$). In the neighborhood of t_ν, we have therefore

$$\alpha(t) = c \log(t - t_\nu) + \sum_{m=-\infty}^{+\infty} c_m (t - t_\nu)^m.$$

On the other hand, since $\alpha(t) = \int^z A$, $z \varepsilon \Gamma_t$, it is clear that the limit

---

[50] Especially if Γ is linearly equivalent to an irreducible curve, then Γ is connected. Cf. Zariski [30], p. 36.

$\lim_{\epsilon\to+0} \alpha(\epsilon e^{i\theta} + t_\nu)$ exists and is independent of θ. Hence, in the above expansion of $\alpha(t)$, c and c_m, $m < 0$, must be zero. Thus $\alpha(t)$ is everywhere holomorphic and therefore $\alpha(t)$ must be a constant. This shows that $\int^z A$ reduces to a constant, proving that A vanishes identically on $\mathfrak{M}$, q. e. d.

Combined with (6.12), the above Theorem 6.3 yields immediately a theorem of F. Enriques[51] to the effect that the *differential deficiency* $dd(\Gamma)$ *of an arbitrary irreducible curve* Γ *on an algebraic surface* $\mathfrak{M}$ *is never greater than the irregularity* q *of* $\mathfrak{M}$ *and there exists on* $\mathfrak{M}$ *an irreducible curve* Γ *for which* $dd(\Gamma) = q$..

In what follows we assume that *there exists on* $\mathfrak{M}$ *at least one canonical divisor* K. Then, using (5.14), we get from Theorem 6.2 the following

THEOREM 6.4. *The dimension of the adjoint class* $\{K + \Gamma\}$ *of* $\{\Gamma\}$ *is given by*

$$\dim\{K + \Gamma\} = g + \pi_\Gamma + m - 1 - q + k. \tag{6.13}$$

This Theorem 6.4 yields immediately *Riemann-Roch's inequality for the adjoint class* $\{\Gamma'\} = \{K + \Gamma\}$ of $\{\Gamma\}$. Defining the *virtual genus* π' of Γ' by

$$2\pi' - 2 = I(\Gamma', \Gamma') + I(K, \Gamma'),$$

we get readily from (6.13) the formula

$$\dim\{\Gamma'\} = I(\Gamma', \Gamma') - \pi' + g - q + 2 + m - 1 + k,$$

which yields immediately *Riemann-Roch's inequality*

$$\dim\{\Gamma'\} \geqq I(\Gamma', \Gamma') - \pi' + g - q + 2.$$

Furthermore we get the following

THEOREM 6.5. *The superabundance of the adjoint class* $\{K + \Gamma\}$ *of the divisor class* $\{\Gamma\}$ *of an arbitrary curve* Γ *is given by*

$$\sup\{K + \Gamma\} = m - 1 + k, \tag{6.14}$$

where m *is the number of conected components of* Γ *and* k *is the number of linearly independent simple Picard integrals* $\int^z A$ *of the first kind which are constant on each connected component of* Γ.

---

[51] Enriques [8]; Goldman [9], § 6. Cf. also Zariski [30], p. 66.

Combining (6.14) with (6.12), we get immediately

$$\sup\{K+\Gamma\} = m-1+q-dd(\Gamma). \tag{6.15}$$

Especially if Γ is a *connected* curve, we have

$$\sup\{K+\Gamma\} = q-dd(\Gamma). \tag{6.16}$$

Hence we get from Theorem 6.3 the following

THEOREM 6.6. *If* Γ *is connected and* $\dim\{\Gamma\} \geqq 2$, *then the superabundance of the adjoint class* $\{K+\Gamma\}$ *of* Γ *is zero*: $\sup\{K+\Gamma\} = 0$.

It is to be noted here that

$$dd(\Gamma) = \operatorname{def}(K+\Gamma/\Gamma) \tag{6.17}$$

whenever $\operatorname{def}(K+\Gamma/\Gamma)$ can be defined. In fact, if there exist a canonical divisor $K = (W)$ and a divisor $D \,\varepsilon\, \{\Gamma\}$ such that K and D have no component in common with Γ, then it can be easily verified in the same manner as in the proof of Theorem 5.4 that the mapping $f \rightarrow \xi = fW_{12}(\phi_p)h_p(D)\sigma_p$ maps $\mathfrak{f}(\Gamma, K+D)$ *onto* $\Xi(\Gamma)$ one-to-one so that $\dim \Xi(\Gamma) = \dim \mathfrak{f}(\Gamma, K+D)$. Hence we get

$$dd(\Gamma) = \dim \mathfrak{f}(\Gamma, K+D) - \dim\{K+D\} + \dim\{K\},$$

proving (6.17).

In case $\mathfrak{M}$ is an algebraic surface, the above Theorem 6.6 is reduced to a special case of a theorem of F. Severi[52] to the effect that, *if an irreducible curve* Γ *on* $\mathfrak{M}$ *is a member of a continuous system which is not an irrational pencil, then the superabundance of the adjoint system* $\{K+\Gamma\}$ *of* $\{\Gamma\}$ *is zero.* This theorem of Severi can be readily deduced from Theorem 6.5. Indeed, as was proved by Severi,[53] if Γ satisfies the hypothesis of the theorem, then there exists on $\mathfrak{M}$ no non-trivial simple Picard integral of the first kind which reduces to a constant on Γ; hence, by Theorem 6.5, $\sup\{K+\Gamma\} = 0$.

7. Algebraic surfaces. In this Section $\mathfrak{M}$ is assumed to be an *algebraic surface* without singularities in the projective space $\mathfrak{C}^N$ of complex dimension N. We denote the projective coordinates in $\mathfrak{C}^N$ by $\zeta_0, \zeta_1, \cdots, \zeta_N$. Then the Kählerian metric $ds^2 = 2\sum g_{\alpha\beta^*}(dz^\alpha d\bar{z}^\beta)$ on $\mathfrak{M}$ is given by

$$g_{\alpha\beta^*} = \partial^2 Z/\partial z^\alpha \partial \bar{z}^\beta, \qquad Z = (1/2\pi)\log\{\sum_{k=0}^{N} |\zeta_k/\zeta|^2\},$$

---

[52] Severi [23]. See also Zariski [30], p. 144.

[53] Severi [22]; Zariski [30], p. 144.

where $\zeta = l_0\zeta_0 + l_1\zeta_1 + \cdots + l_N\zeta_N$ is an arbitrary *fixed* linear form of $\zeta_0, \zeta_1, \cdots, \zeta_N$ (it can be readily verified that $g_{\alpha\beta^*}$ are independent of the choice of ζ). The hyperplane section of $\mathfrak{M}$ cut out by the hyperplane $\zeta = 0$ will be denoted by E. We choose the linear form ζ so that E and Γ have only *simple intersection points*. Now, let us consider the quantity $Q_f[\Omega \cdot A]$ appearing in the condition (3.32). Putting

$$\kappa = \sum \kappa_\alpha \cdot dz^\alpha, \qquad \kappa_\alpha = \partial Z/\partial z^\alpha,$$

we get $\Omega = -id\kappa$ and therefore

$$i\Omega \cdot A = d(\kappa \cdot A) = \sum_\alpha \partial(\kappa_1 A_2 - \kappa_2 A_1)/\partial \bar{z}^\alpha \cdot dz^1 dz^2 d\bar{z}^\alpha.$$

Hence, considering $(\kappa_1 A_2 - \kappa_2 A_1) f h_p \sigma_p$ as a 1-form on $\mathbf{\Gamma}$, we have

$$iQ_f[\Omega \cdot A] = \int_{\mathbf{\Gamma}} \sum_\alpha \partial(\kappa_1 A_2 - \kappa_2 A_1)/\partial \bar{z}^\alpha \cdot f h_p \sigma_p d\bar{\phi}_p{}^\alpha$$
$$= -\int_{\mathbf{\Gamma}} d\{(\kappa_1 A_2 - \kappa_2 A_1) f h_p \sigma_p\}.$$

The 1-form $(\kappa_1 A_2 - \kappa_2 A_1) f h_p \sigma_p$ is of class C^∞ on $\mathbf{\Gamma}$ except for the points p on $|\mathfrak{c}|$ and on $|E \cdot \Gamma|$, while, since, by our assumption, E and Γ have only simple intersection points, $|\mathfrak{c}|$ and $|E \cdot \Gamma|$ have no common point. We have therefore

$$iQ_f[\Omega \cdot A] = \Big(\sum_{p \varepsilon |\mathfrak{c}|} + \sum_{p \varepsilon |E \cdot \Gamma|}\Big) \lim_{\epsilon \to +0} \int_{|\tau_p| = \epsilon} (\kappa_1 A_2 - \kappa_2 A_1) f h_p \sigma_p,$$

while, using Lemma 3.1, one can readily deduce from the condition (β) for f the identity

$$\sum_{\phi(p) = \mathfrak{p}} \lim_{\epsilon \to +0} \int_{|\tau_p| = \epsilon} (\kappa_1 A_2 - \kappa_2 A_1) f h_p \sigma_p = 0$$

for each singular point $\mathfrak{p}$ of Γ. Hence we get

$$iQ_f[\Omega \cdot A] = \sum_{p \varepsilon |E \cdot \Gamma|} \lim_{\epsilon \to +0} \int_{|\tau_p| = \epsilon} (\kappa_1 A_2 - \kappa_2 A_1) f h_p \sigma_p. \tag{7.1}$$

Consider an arbitrary point $p \,\varepsilon\, |E \cdot \Gamma|$. Assuming that $\mathfrak{p} = \phi(p)$ is not on the hyperplane $\zeta_0 = 0$, we put $\zeta_0 = 1$ in $\mathfrak{U}_\mathfrak{p}$. Then the inhomogeneous coordinates $\zeta_1, \zeta_2, \cdots, \zeta_N$ of the point $z = (z^1, z^2)$ in $\mathfrak{U}_\mathfrak{p}$ are holomorphic functions of z^1, z^2 and therefore

$$\kappa = -(1/2\pi)\, d \log \zeta + \text{continuous form}, \qquad \text{in } \mathfrak{U}_\mathfrak{p}.$$

Choosing the local coordinates z^1, z^2 so that $R_p = z^1$, $\zeta = z^2$ in $\mathfrak{U}_\mathfrak{p}$ (this is possible since $\mathfrak{p}$ is assumed to be a simple intersection point of E and Γ), we get therefore

$$\lim_{\epsilon\to+0}\int_{|\tau_p|=\epsilon}(\kappa_1 A_2 - \kappa_2 A_1) f h_p \sigma_p = i A_1(\mathfrak{p}) f h_p(p),$$

where $fh_p(p)$ denotes the value of the holomorphic function fh_p at p. A_1 can be interpreted as the coefficient of the differential $A_E = A_1 dz^1$ induced on E by A and represented as

$$A_1(\mathfrak{p}) = \mathrm{Res}_{\mathfrak{p}}[A_E/R_{\mathfrak{p}}] = (1/2\pi i)\oint_{\mathfrak{p}} R_{\mathfrak{p}}^{-1} A_E,$$

where $\oint_{\mathfrak{p}}$ denotes the integral extended over the small circle *on* E with the center $\mathfrak{p}$. Thus we get

$$\lim_{\epsilon\to+0}\int_{|\tau_p|=\epsilon}(\kappa_1 A_2 - \kappa_2 A_1) f h_p \sigma_p = i f h_p(p)\,\mathrm{Res}_{\mathfrak{p}}[A_E/R_p].$$

Combined with (7.1), this yields the *important formula*

$$Q_f[\Omega\cdot A] = \sum_{p\,\varepsilon\,|E\cdot\Gamma|} f h_p(p)\,\mathrm{Res}_{\phi(p)}[A_E/R_p], \tag{7.2}$$

where the expression $fh_p(p)\mathrm{Res}_{\phi(p)}[A_E/R_p]$ *is independent of the choice of the local equation* $R_p = 0$.

Now, assume that Γ is *an irreducible curve without singular points* and consider an arbitrary divisor D not containing Γ as one of its components. In this case we can consider that $\boldsymbol{\Gamma} = \Gamma$, so that $\mathfrak{f}(\Gamma, D)$ is simply the space consisting of all meromorphic functions f on Γ which are multiples of $-D\cdot\Gamma$. The deficiency[54] of D *on* Γ is given therefore by

$$\mathrm{def}(D/\Gamma) = \dim_{\Gamma}\{D\cdot\Gamma\} - \dim\{D\} + \dim\{D-\Gamma\},$$

where $\dim_{\Gamma}\{D\cdot\Gamma\}$ denotes the dimension of the divisor class $\{D\cdot\Gamma\}$ on the curve Γ. Now, assuming that $\Gamma - D$ is linearly equivalent to a (reducible or irreducible) curve Δ which does not contain Γ as one of its components and *is connected in* $\mathfrak{M}-\Gamma$, we shall determine $\mathrm{def}(D/\Gamma)$. Let F_0 be a meromorphic function on $\mathfrak{M}$ such that $(F_0) = D + \Delta - \Gamma$. We choose the meromorphic functions $R_{\mathfrak{p}}(D; z)$, $R_{\mathfrak{p}}(\Delta; z)$ expressing respectively the divisors D, Δ in $\mathfrak{U}_{\mathfrak{p}}$ so that $F_0 = R_{\mathfrak{p}}(D)R_{\mathfrak{p}}(\Delta)/R_{\mathfrak{p}}$. Then, as was shown in Section **5**, one gets a system $\{h_{\mathfrak{p}}\}$ defining the characteristic divisor $\mathfrak{d}$ by putting $h_{\mathfrak{p}} = R_{\mathfrak{p}}(D; \phi_{\mathfrak{p}})R_{\mathfrak{p}}(\Delta; \phi_{\mathfrak{p}})$ (since $\boldsymbol{\Gamma} = \Gamma$, it is not necessary to distinguish p from $\mathfrak{p} = \phi(p)$). Denote by M an arbitrary meromorphic function in

---

[54] Goldman [9], § 6.

$\mathfrak{F}(D+\Delta)$. Then the mapping $M \rightarrow F = MF_0$ maps $\mathfrak{F}(D+\Delta)$ *onto* $\mathfrak{F}(\Gamma)$ one-to-one. Again, by virtue of Theorem 3.4, every $F \,\varepsilon\, \mathfrak{F}(\Gamma)$ corresponds one-to-one up to an additive constant to $f \,\varepsilon\, \mathfrak{f}(\Gamma, \mathfrak{d})$ satisfying $Q_f[\Omega \cdot A_\lambda] = 0$ $(\lambda = 1, 2, \cdots, q)$ by means of the relation $f = R_\mathfrak{p}F/h_\mathfrak{p}$, while, since $h_\mathfrak{p}/R_\mathfrak{p} = R_\mathfrak{p}(D)R_\mathfrak{p}(\Delta)/R_\mathfrak{p} = F_0$, the relation $f = R_\mathfrak{p}F/h_\mathfrak{p}$ is equivalent to $f(\tau_\mathfrak{p}) = M(\phi_\mathfrak{p}(\tau_\mathfrak{p}))$. Hence, every meromorphic function $M \,\varepsilon\, \mathfrak{F}(D+\Delta)$ corresponds one-to-one up to an additive term cF_0^{-1} (c being an arbitrary constant) to a meromorphic function $f \,\varepsilon\, \mathfrak{f}(\Gamma, \mathfrak{d})$ satisfying $Q_f[\Omega \cdot A_\lambda] = 0$ $(\lambda = 1, 2, \cdots, q)$ by means of the relation $f = M(\phi_\mathfrak{p})$. Obviously $\mathfrak{d} = D \cdot \Gamma + \Delta \cdot \Gamma$ and therefore $\mathfrak{f}(\Gamma, \mathfrak{d}) = \mathfrak{f}(\Gamma, D + \Delta)$. Now, if $M \,\varepsilon\, \mathfrak{F}(D)$, then $f = M(\phi_\mathfrak{p})$ belongs clearly to the subspace $\mathfrak{f}(\Gamma, D)$ of $\mathfrak{f}(\Gamma, D+\Delta)$; moreover any $M \,\varepsilon\, \mathfrak{F}(D)$ is determined uniquely by $f = M(\phi_\mathfrak{p})$, since F_0^{-1} does not belong to $\mathfrak{F}(D)$. Conversely, *if* $f = M(\phi_\mathfrak{p})$ *belongs to* $\mathfrak{f}(\Gamma, D)$, *then we can find* $M_1 \,\varepsilon\, \mathfrak{F}(D)$ *such that* $M_1(\phi_\mathfrak{p}) = f$. To prove this, we shall first show that, if $f = R_\mathfrak{p}F/h_\mathfrak{p}$ belongs to $\mathfrak{f}(\Gamma, D)$, then the function F_Δ induced on Δ by F is holomorphic everywhere on Δ. Clearly F_Δ is holomorphic on Δ except for the points on $|\Delta \cdot \Gamma|$. Now, let $\mathfrak{p}$ be a point on $|\Delta \cdot \Gamma|$. Then since Γ is a curve without singular points, we can choose the local coordinates z^1, z^2 in $\mathfrak{U}_\mathfrak{p}$ so that $R_\mathfrak{p} = z^1$ and that $z^2 = \tau_\mathfrak{p}$ on Γ. Put

$$N(z^1, z^2) = z^1 F = MR_\mathfrak{p}(D)R_\mathfrak{p}(\Delta).$$

Then, since $f \,\varepsilon\, \mathfrak{f}(\Gamma, D)$ implies that $M(0, z^2)R_\mathfrak{p}(D; 0, z^2) = f(z^2)R_\mathfrak{p}(D; 0, z^2)$ is holomorphic in z^2, we have

$$N(0, z^2) \equiv 0(R_\mathfrak{p}(\Delta; 0, z^2)).$$

This shows that

$$N(z^1, z^2) \equiv 0(z^1, R_\mathfrak{p}(\Delta; z^1, z^2)).$$

Now, let $\Delta^{(1)}$ be any irreducible branch of Δ passing through $\mathfrak{p}$ and $z^\alpha = \psi^\alpha(\tau_\Delta)$ $(\alpha = 1, 2)$ be a parametric representation of $\Delta^{(1)}$. Then we get from the above result

$$N(\psi^1(\tau_\Delta), \psi^2(\tau_\Delta)) \equiv 0(\psi^1(\tau_\Delta)),$$

proving that $F_\Delta = F(\psi(\tau_\Delta)) = N(\psi(\tau_\Delta))/\psi^1(\tau_\Delta)$ is a holomorphic function of the local uniformization variable τ_Δ on $\Delta^{(1)}$. Thus F_Δ is holomorphic everywhere on Δ and therefore F_Δ must be constant on each irreducible component of Δ, while, since by hypothesis Δ is connected in $\mathfrak{M} - \Gamma$, any two irreducible components of Δ have at least one common point in $\mathfrak{M} - \Gamma$. Hence F_Δ reduces to a constant on $\Delta : F_\Delta = c$. Now, putting $F_1 = F - c$, we get a meromorphic function F_1 satisfying $(F_1) \geqq \Delta - \Gamma$ and $R_\mathfrak{p}F_1/h_\mathfrak{p} = f$. The corresponding meromorphic function $M_1 = F_1/F_0$ belongs clearly to $\mathfrak{F}(D)$ and satisfies $M_1(\phi_{\mathfrak{p}}) = f$. Thus *the mapping* $M \to f = M(\phi_{\mathfrak{p}})$ *gives the one-to-one correspondence between meromorphic functions* $M \varepsilon \mathfrak{F}(D)$ *and meromorphic functions* $f \varepsilon \mathfrak{f}(\Gamma, D)$ *satisfying* $Q_f[\Omega \cdot A_\lambda] = 0$ $(\lambda = 1, 2, \cdots, q)$.

Let j_D be the number of linearly independent simple differentials A of the first kind satisfying $Q_f[\Omega \cdot A] = 0$ for all $f \varepsilon \mathfrak{f}(\Gamma, D)$. Then the number of linearly independent conditions imposed on $f \varepsilon \mathfrak{f}(\Gamma, D)$ by $Q_f[\Omega \cdot A_\lambda] = 0$ $(\lambda = 1, 2, \cdots, q)$ is $q - j_D$ and therefore, by virtue of the above results,

$$\operatorname{def}(D/\Gamma) = \dim \mathfrak{f}(\Gamma, D) - \dim \mathfrak{F}(D) = q - j_D.$$

Combined with (7.2), this proves the following

THEOREM 7.1. *Assume* Γ *to be an irreducible curve without singular point. Let* D *be an arbitrary divisor not containing* Γ *as its component such that* $\Gamma - D$ *is linearly equivalent to a curve* Δ *which does not contain* Γ *as its component and is connected in* $\mathfrak{M} - \Gamma$, *and let* j_D *be the number of linearly independent simple differentials* A *of the first kind satisfying*

$$\sum_{\mathfrak{p} \varepsilon |E \cdot \Gamma|} f h_{\mathfrak{p}}(\mathfrak{p}) \operatorname{Res}_{\mathfrak{p}}[A_E/R_{\mathfrak{p}}] = 0, \qquad \text{for all } f \varepsilon \mathfrak{f}(\Gamma, D),$$

where E *is an arbitrary hyperplane section of* $\mathfrak{M}$ *cutting out on* Γ *only simple intersection points. Then we have*

$$(7.3) \qquad \operatorname{def}(D/\Gamma) = q - j_D.$$

From this theorem it follows immediately the following

THEOREM 7.2. *Let* Γ *be an irreducible curve without singular points and* D *be an arbitrary divisor not containing* Γ *as its component. If* $\Gamma - D - E$ *is linearly equivalent to a curve* Δ *not containing* Γ *as its component, then we have* $\operatorname{def}(D/\Gamma) = 0$.

Proof. We choose E so that E cuts out on Γ only simple intersection points which do not lie on $|D|$ nor on $|\Delta|$ and that $\Delta + E$ is connected in $\mathfrak{M} - \Gamma$. Since Γ is linearly equivalent to $D + \Delta + E$, we may assume that

$$(7.4) \qquad h_{\mathfrak{p}} = R_{\mathfrak{p}}(D; \phi_{\mathfrak{p}})\ R_{\mathfrak{p}}(\Delta; \phi_{\mathfrak{p}})\ R_{\mathfrak{p}}(E; \phi_{\mathfrak{p}}).$$

Now, for each $\mathfrak{p} \varepsilon |E \cdot \Gamma|$, every $f \varepsilon \mathfrak{f}(\Gamma, D)$ is holomorphic at $\mathfrak{p}$, while (7.4) shows that $h_{\mathfrak{p}}(\mathfrak{p}) = 0$. Hence we have, for arbitrary A,

$$\sum_{\mathfrak{p} \varepsilon |E \cdot \Gamma|} f h_{\mathfrak{p}}(\mathfrak{p}) \operatorname{Res}_{\mathfrak{p}}[A_E/R_{\mathfrak{p}}] = 0, \qquad \text{for all } f \varepsilon \mathfrak{f}(\Gamma, D).$$

Consequently $j_D = q$ and therefore, by (7.3), $\operatorname{def}(D/\Gamma)$ is zero.

For an arbitrary given divisor D, the divisor class $\{D+nE\}$ contains an irreducible curve without singularities if the positive integer n is chosen sufficiently large.[55] Using this fact, we can deduce from Theorem 7.2 immediately the following

THEOREM 7.3 (O. Goldman[56]). *Let L, D be two arbitrary given divisors and Γ_n be an irreducible curve without singularities belonging to the divisor class $\{L+nE\}$, E being the hyperplane section. Then, if n is sufficiently large, we have* $\operatorname{def}(D/\Gamma_n)=0$.

The *superabundance* of an arbitrary divisor class $\{D\}$ is, by definition,

$$\sup\{D\}=\dim\{D\}+\dim\{K-D\}-I(D,D)+\pi_D-g+q-2,$$

where π_D is the *virtual genus* of D defined by $2\pi_D-2=I(D,D)+I(K,D)$. The *theorem of Riemann-Roch* on algebraic surfaces asserts that $\sup\{D\}$ *is always non-negative.* The divisor class $\{D\}$ is called *regular* if $\sup\{D\}$ and $\dim\{K-D\}$ are both zero.

Let D is an arbitrary divisor. Then, if n is sufficiently large, $\{D+nE-K\}$ is a divisor class of an irreducible curve Δ and therefore $\{D+nE\}=\{K+\Delta\}$; moreover, it is obvious that $\dim\{\Delta\}\geqq 2$ for sufficiently large n. Hence, by Theorem 6.6, we get the following

THEOREM 7.4 (G. Castelnuovo[57]). *For an arbitrary given divisor D, the divisor class $\{D+nE\}$ is regular if n is sufficiently large.*

Choose n so large that $\{D+nE\}$ is regular and contains a curve Γ. Then, using (5.17), we get immediately $chd\{\Gamma\}=q$. Combined with (3.35), this proves a theorem of G. Castelnuovo[58] to the effect that *the characteristic deficiency of a curve is always not greater than the irregularity q of the surface and there exists a curve having the characteristic deficiency q.* Again, combined with Theorem 3.4, the above result yields the following

THEOREM 7.5.[59] *Let Γ_n be a curve in the divisor class $\{D+nE\}$, D being an arbitrary fixed divisor. Then, if n is sufficiently large, there exists on $\mathfrak{M}$ exactly $2q$ independent simple Picard integrals of the second kind which are multiples of $-\Gamma_n$.*

---

[55] See Goldman [9], § 5.

[56] Goldman [9], § 6.

[57] Castelnuovo [4]; Goldman [9], § 7.

[58] Castelnuovo [4]; see also Zariski [30], pp. 69-70.

[59] Cf. Zariski [30], pp. 122-128.

As was shown by O. Goldman,[60] *the theorem of Riemann-Roch on algebraic surfaces can be readily deduced from Theorem 7.3 and Theorem 7.4.* We shall reproduce here his deduction briefly. Let D be a given divisor and Γ be an irreducible curve without singularities belonging to the divisor class $\{nE\}$, n being a sufficiently large positive integer. Then we have, by Theorem 7.3, $\operatorname{def}(K - D/\Gamma) = 0$ and therefore

$$\dim \{K - D\} = \dim_\Gamma \{K \cdot \Gamma - D \cdot \Gamma\},$$

where $\dim_\Gamma \{K \cdot \Gamma - D \cdot \Gamma\}$ denotes the dimension of the intersection divisor $K \cdot \Gamma - D \cdot \Gamma$ on the curve Γ. Again, since $\Gamma \cdot \Gamma + K \cdot \Gamma$ is a canonical divisor on Γ, we have, by the theorem of Riemann-Roch on the curve,

$$\dim_\Gamma \{D \cdot \Gamma + \Gamma \cdot \Gamma\} = I(\Gamma, D + \Gamma) - \pi_\Gamma + \dim_\Gamma \{K \cdot \Gamma - D \cdot \Gamma\} + 1,$$

while, it is obvious that

$$\dim_\Gamma \{D \cdot \Gamma + \Gamma \cdot \Gamma\} \geqq \dim \{D + \Gamma\} - \dim \{D\}.$$

Consequently we get

$$\dim \{D\} + \dim \{K - D\} \geqq \dim \{D + \Gamma\} - I(\Gamma, D + \Gamma) + \pi_\Gamma - 1.$$

On the other hand, Theorem 7.4 shows that

$$\dim \{D + \Gamma\} = I(D + \Gamma, D + \Gamma) - \pi_{D+\Gamma} + g - q + 2.$$

Hence we get

$$\dim \{D\} + \dim \{K - D\} \geqq I(D, D + \Gamma) - \pi_{D+\Gamma} + \pi_\Gamma + g - q + 1,$$

yielding immeditely

$$\dim \{D\} + \dim \{K - D\} \geqq I(D, D) - \pi_D + g - q + 2,$$

q. e. d.

Remark. The arithmetic genus of an algebraic surface $\mathfrak{M}$ is, by definition, the constant p_a appearing in the "postulation formula"

$$\dim \{nE\} = c_0 n^2 + c_1 n + p_a + 1, \qquad \text{(for large } n\text{)},$$

where E denotes a hyperplane section of $\mathfrak{M}$ and c_0, c_1 are the constants not depending on n. On the other hand, it follows from Theorem 7.4 that, for sufficiently large n, $\dim \{nE\}$ is given by

$$\dim \{nE\} = \tfrac{1}{2} I(E, E) n^2 - \tfrac{1}{2} I(K, E) n + g - q + 1.$$

---

[60] Goldman [9], § 7.

Comparing this with the postulation formula, we infer immediately that *the difference* $g - p_a$ *between the geometric genus* g *and the arithmetic genus* p_a *of an algebraic surface* $\mathfrak{M}$ *is equal to the number* q *of linearly independent simple differentials of the first kind attached to* $\mathfrak{M}$.[61]

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] H. Behnke and P. Thullen, "Theorie der Funktionen Mehrerer Komplexer Veränderlichen," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 3, No. 3, Berlin (1934).

[2] S. Bochner and W. T. Martin, "Several complex variables," Princeton (1948).

[3] H. Cartan, "Espaces fibrés analytiques complexes," *Séminaire Bourbaki* (Décembre 1950).

[4] G. Castelnuovo, "Alcune proprietà fondamentali dei sistemi lineari di curve tracciati sopra una superficie algebrica," *Annali di Mathematica, pura et applicata*, Serie II, Tomo 25 (1897), pp. 235-318.

[5] S. Chern, "Some new viewpoints in differential geometry in the large," *Bulletin of the American Mathematical Society*, vol. 52 (1946), pp. 1-30.

[6] ———, "Characteristic classes of Hermitian manifolds," *Annals of Mathematics*, vol. 47 (1946), pp. 85-121.

[7] B. Eckmann and H. Guggenheimer, "Formes differentielles et métrique hermitienne sans torsion, I. Structure complexe, formes pures, II. Formes de class k, formes analytiques," *Comptes Rendus*, Tome 299 (1950), pp. 464-466, pp. 489-491.

[8] F. Enriques, "Introduzione alla geometria sopra le superficie algebriche," *Memorie della Società Italiana delle Scienze* Dtta Dei 40, Serie III, vol. 10 (1896).

[9] O. Goldman, "On the theory of algebraic surfaces," to appear in the Memoirs of the American Mathematical Society.

[10] D. Gorenstein, "An arithmetic theory of adjoint plane curves," unpublished.

[11] F. Hartogs, "Über die aus den singulären Stellen einer analytischen Funktion mehrerer Veränderlichen bestehenden Gebilde," *Acta Mathematica*, vol. 32 (1909), pp. 57-79.

[12] W. V. D. Hodge, "The theory and applications of harmonic integrals," Cambridge University Press (1941).

[13] E. Kähler, "Über eine bemerkenswerte Hermitishe Metrik," *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität*, vol. 9 (1933), pp. 173-186.

[14] K. Kodaira, "Harmonic fields in Riemannian manifolds," *Annals of Mathematics*, vol. 50 (1949), pp. 587-665.

---

[61] See Zariski [30], pp. 122-128.

[15] S. Lefschetz, "L'Analysis situs et la géometrie algébrique," Paris (1924).

[16] E. Picard and G. Simart, "Théorie des fonctions algébriques de deux variables indépendantes," 2 volumes, Paris (1897-1906).

[17] H. Poincaré, "Sur les résidues des intégrales doubles," *Acta Mathematica*, vol. 9 (1886-1887), pp. 321-380.

[18] G. de Rham, "Sur la théorie des formes différentielles harmoniques," *Annales de l'Université de Grenoble*, vol. 22 (1946), pp. 135-152.

[19] ——— and P. Bidal, "Les formes différentielles harmoniques," *Commentarii Mathematici Helvetici*, vol. 19 (1946), pp. 1-49.

[20] ——— and K. Kodaira, "Harmonic integrals," (Mimeographed notes), Institute for Advanced Study, Princeton (1950).

[21] L. Schwartz, "Theorie des distributions," vol. I, II, Paris (1950-1951).

[22] F. Severi, "Intorno al teorema d'Abel sulle superficie algebriche ed alla riduzione a forma normale degl'integrali di Picard," *Rendiconti del Circolo Mathematico di Palermo*, vol. 21 (1906), pp. 257-282.

[23] ———, "Sulla regolarità del sistema aggiunto ad un sistema lineare di curve appartenente ad una superficie algebrica," *Atti della Reale Accademia Nazionale dei Lincei*, Serie V, vol. 17 (1908).

[24] A. Weil, "Sur la théorie des formes différentielles attachées à une variété analytique complexe," *Commentarii Mathematici Helvetici*, vol. 20 (1947), pp. 110-116.

[25] ———, "Fibre-spaces in algebraic geometry," (Mimeographed notes), Conference on algebraic geometry and algebraic number theory, Chicago (1949).

[26] ———, "Foundations of algebraic geometry," *American Mathematical Society Colloquium Publications*, vol. 29 (1946).

[27] H. Weyl, "Die Idee der Riemannschen Flache," Berlin (1913).

[28] ———, "Method of orthogonal projection in potential theory," *Duke Mathematical Journal*, vol. 7 (1940), pp. 411-444.

[29] ———, "The classical groups," Princeton (1946).

[30] O. Zariski, "Algebraic surfaces," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 3, no. 5 (1935).

GAUSSIAN CURVATURE AND LOCAL EMBEDDING.*

By PHILIP HARTMAN ** and AUREL WINTNER

1. The embedding theorems. As recently shown in [2], [3], [4], the usual assumptions of differentiability in some standard theorems of differential geometry can be lightened considerably if considerations involving the Gauss formula (and its variants) for the surface curvature K are replaced by others involving the "curvatura integra," $\int\int K d\sigma$; cf. [8], pp. 43-44, [6], p. 135 and § 2 below. The present paper develops these ideas farther, with applications to the embedding problem.

A function of one or several variables will be called of class $C^n(\lambda)$, where $n = 0, 1, \cdots$ and $0 < \lambda < 1$, on a domain if the function is of class C^n on that domain (that is, if it possesses there continuous n-th partial derivatives) and if all the n-th derivatives satisfy a uniform Hölder condition of order λ (with respect to all of the variables, and on every compact subset of the domain). Let $(g_{ik}) = (g_{ik}(u, v))$, where $i, k = 1, 2$, be a symmetric positive-definite matrix defined on some (u, v)-domain. The line element

$$(1) \qquad ds^2 = g_{ik}du^i du^k, \qquad \text{where } (u^1, u^2) = (u, v),$$

will be said to be of class C^n or of class $C^n(\lambda)$ if each coefficient function in (1) belongs to the class C^n or $C^n(\lambda)$, respectively. For $n \geqq 1$, the (g_{ik}) or (1) will be said to possess a local embedding of class C^n or $C^n(\lambda)$ at a given point (u_0, v_0), if, in a vicinity of (u_0, v_0), there exist three functions $x = x(u, v)$, $y = y(u, v)$, $z = z(u, v)$ of class C^n or $C^n(\lambda)$, respectively, satisfying

$$(2) \qquad dx^2 + dy^2 + dz^2 = g_{ik}du^i du^k;$$

that is, if there exists, in the 3-dimensional Euclidean space, a 2-dimensional surface, of class C^n or $C^n(\lambda)$, on which (1) is the squared element of arc-length.

Several extensions of the theorem of [2], p. 563, will be proved. In the elliptic case, the result will be as follows:

(I) *Let* $0 < \mu < \lambda < 1$. *In a neighborhood of* $(u, v) = (0, 0)$, *let* (1)

---

* Received December 16, 1950.

** John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

be a positive-definite line element of class $C^1(\lambda)$ *which possesses a positive curvature* $K = K(u, v)$ *of class* $C^0(\lambda)$. *Then* (1) *has a local embedding of class* $C^2(\mu)$.

The analogue of (I), in which $C^0(\lambda)$, $C^1(\lambda)$, $C^2(\mu)$ are replaced by $C^{n-1}(\lambda)$, $C^n(\lambda)$, $C^{n+1}(\mu)$, respectively, where $n \geqq 2$, was proved in [2]. The definition of the curvature K of (1), when (1) is of class C^1, will be given in § 2 below, following the ideas mentioned at the beginning of § 1. Whether or not the $C^2(\mu)$ in the assertion of (I) can be improved to $C^2(\lambda)$ will remain undecided. Also, it will be left open whether or not the Hölder conditions can be dropped in both the assumption and assertion of (I). It will, however, be proved that they can be omitted in the hyperbolic analogue of (I):

(II) *In a neighborhood of* $(u, v) = (0, 0)$, *let* (1) *be a positive-definite line element, of class* C^1, *which possesses a continuous negative curvature* $K = K(u, v)$. *Then* (1) *has a local embedding of class* C^2.

Since the surface (x, y, z) of class C^2 supplied by (II) has a continuous Gaussian curvature which is identical with K, the assumption in (II) that (1) possess a continuous curvature is necessary. In this regard, cf. the remarks in [2], p. 554 or the footnote in [4], p. 144.

Corresponding to (II), the Hölder conditions can be omitted in the hyperbolic case of the theorem of [2], p. 563, as follows:

(III) *Let* $n \geqq 1$. *In a neighborhood of* $(u, v) = (0, 0)$, *let* (1) *be a positive-definite line element, of class* C^n, *which posseses a negative curvature of class* C^{n-1}. *Then* (1) *has a local embedding of class* C^{n+1}.

It remains undecided whether the assertion of (III) does or does not remain true if the curvature, instead of being negative, is positive.

2. The generalized definition of curvature.

Let (1) be a positive-definite line element on a simply connected (u, v)-domain R. If (1) is of class C^2, the curvature $K = K(u, v)$ of (1) can be defined by the standard formula

$$K = (B_u - A_v)/g, \tag{3}$$

where $g = \det^{\frac{1}{2}}(g_{ik}) > 0$ and

$$A = \tfrac{1}{2}g^{-1}[(g_{12}/g_{11})g_{11u} + g_{11v} - g_{12u}], \tag{4}$$

$$B = \tfrac{1}{2}g^{-1}[(g_{12}/g_{11})g_{11v} - g_{22u} + g_{12v}]; \tag{5}$$

cf., e. g., [1], p. 216. If J is a Jordan curve in R and is piecewise of class C^1, and if T is the interior of J, then (3) and Green's theorem imply that

$$\phi(T) = \int_J A du + B dv, \tag{6}$$

where the set-function $\phi(T)$ is the "curvatura integra,"

$$\phi(T) = \iint_T K d\sigma, \qquad (d\sigma = g\, du\, dv). \tag{7}$$

In what follows, by "an admissible Jordan domain T" will be meant a domain T of the type admitted in (6) and (7). The formula (6) will be used as a basis for the following definition:

Definition. If (1) is a positive-definite line element of class C^1 (not necessarily C^2) on a simply-connected domain R, the "curvatura integra," $\phi(T)$, of (g_{ik}) or (1) is defined for admissible Jordan domains T by (6), where A, B are given by (4), (5). If, in addition, the set-function $\phi(T)$ is absolutely continuous, that is, if $\phi(T)$ has a representation of the form (7) in terms of a function $K = K(u, v)$ which is independent of T and which is summable (i. e., class L_1) over every compact subset of R, then (g_{ik}) or (1) will be said to possess the *curvature* $K = K(u, v)$.

In this definition, the function $K(u, v)$, if it exists, is only determined almost everywhere (that is, up to a 2-dimensional zero set in R). But even such a determination of a curvature can be useful; cf., e. g., the proof of Theorem III in [4] (where the assumptions imply, however, that K exists and is bounded almost everywhere on compact subsets of R).

The following standard conventions concerning K will be made below: If there exists a continuous function $K = K(u, v)$ satisfying (6) and (7), then the curvature (1) will be considered to be defined everywhere (instead of almost everywhere) as the unique continuous K; in which case (1) will be said to possess a curvature of class C^0. Corresponding remarks apply to a curvature of class C^n or $C^n(\lambda)$.

Similar extensions of the classical definition of the curvatura integra or of the curvature have been given by replacing the classical formula (3) by the formula of Gauss-Bonnet ([8], pp. 43-44; also [6], p. 135) or by the Frobenius formula for K ([2], p. 562; [3], p. 759). These interpretations are equivalent, as seen from the case $\Gamma \equiv \Gamma^m \equiv 1$ of theorem (i) below (they are clearly equivalent for "smooth" line elements).

It may be mentioned that the equivalence of the two definitions of the curvatura integra, those based on (3) and the formula of Gauss-Bonnet, respectively, leads to a simple geometrical interpretation of the set-function $\phi(T)$, even for line elements (1) which are just of class C^1. In fact, $\phi(T)$ is just the oriented variation of the direction of a vector transported parallel to itself (Levi-Civita) along the boundary, J, of T. In order to see this, it is sufficient to write the formula of Gauss-Bonnet in the form

$$\Sigma\alpha + \Sigma\Delta\theta + \phi(T) = 2\pi,$$

where $\Sigma\alpha$ is the sum of the oriented exterior angles α (measured in the metric (1)) at the vertices of J, the sum $\Sigma\Delta\theta$ is extended over the sub-arcs of J, all of which are supposed to be of class C^1, and $\Delta\theta$ denotes the oriented variation of a continuous determination of θ along one of the sub-arcs, finally $-\theta$ is the oriented angle from the tangent vector of the sub-arc to a vector transported by parallelism along the sub-arc; cf. the proof of the Gauss-Bonnet formula in [6], p. 134, and, for the above C^1-formulation, a remark in [5], § 15.

In applications of the set-function $\phi(T)$, the following fact will be needed:

(i) *Let the function* $\Gamma = \Gamma(u, v)$ *and the positive-definite line element* (1) *be of class* C^1 *on a simply-connected domain* R. *For* $m = 1, 2, \cdots$, *let the functions* $\Gamma^m = \Gamma^m(u, v)$ *and the positive-definite line elements*

$$(1^m) \qquad ds^2 = g^m{}_{ik}du^i du^k \qquad (u^1 = u, u^2 = v)$$

be of class C^2 *on* R *and have the property that, as* $m \to \infty$, *the functions* Γ^m, $g^m{}_{ik}$ *and their first order partial derivatives tend to the functions* Γ, g_{ik} *and their first order partial derivatives, respectively, uniformly on every compact subset of* R. *Let* T *be an admissible Jordan domain in* R. *Then*

$$(8) \qquad \lim_{m\to\infty} \iint_T \Gamma^m K^m d\sigma^m$$

exists and has the value

$$(9) \qquad \int_J \Gamma(A du + B dv) + \iint_T (A\Gamma_v - B\Gamma_u)\, du dv.$$

If, in addition, (1) *has a curvature* K, *then the limit in* (8) *has the value*

$$(10) \qquad \int_T \Gamma K d\sigma.$$

As a consequence of (8), (9) for the case $\Gamma \equiv \Gamma^m \equiv 1$, the relation

$$\lim_{m\to\infty} \phi^m(T) = \phi(T) \tag{11}$$

holds for every admissible Jordan domain T in R.

If $f(u)$, $f^m(u)$ are positive functions of class C^2, say for $-1 \leqq u \leqq 1$, such that, as $m \to \infty$, both $f^m \to f$ and $f^m{}_u \to f_u$ hold uniformly in u, while $f^m{}_{uu} \to f_{uu}$ does not hold, then it is seen, by considering the cases $ds^2 = du^2 + (f(u))^2 dv$ and $ds^2 = du^2 + (f^m(u))^2 dv^2$ of (1) and (1^m), respectively, that $K^m \to K$ need not hold (even if K is continuous). Thus a "direct" verification of the equality of the integral (10) and the limit (8) is not possible.

Proof of (i). It is clear from (6), (7) and Green's theorem that the integral in (8) is equal to the expression (9^m), if (9^m) belongs to Γ^m and (1^m) as (9) does to Γ and (1). Thus it is clear that the limit (8) exists and has the value (9).

If $K = K(u, v)$ is continuous, instead of being merely summable, the equality of the expressions (9) and (10) is an immediate consequence of the lemma in [3], p. 761, since (6) and (7) hold for arbitrary admissible Jordan domains T. However, it is easy to see that the proof of that lemma in [3] can be modified so as to apply to the case of a summable K. Thus, (i) is proved.

(ii) *Let* (1) *be a positive-definite line element of class* C^1 *on a simply-connected domain* R, *and let* (1) *possess a curvature* $K = K(u, v)$. *Let a function* $z = z(u, v)$ *of class* C^2 *on* R *be such that the line element defined by*

$$\gamma_{ik} du^i du^k = g_{ik} du^i du^k - (dz)^2 \tag{12}$$

is positive-definite on R. *Then* (12) *possesses a curvature* $\kappa = \kappa(u, v)$ *satisfying*

$$\kappa\gamma^4 = -g(rs - t^2) + L_1 r + L_2 s + L_3 t + Q + Kg^4. \tag{13}$$

Here γ *and* p, q, r, s, t *denote* $\det^{\frac{1}{2}}(\gamma_{ik}) > 0$ *and the partial derivatives* z_u, z_v, z_{uu}, z_{uv}, z_{vv} *of* z, *respectively*; L_1, L_2, L_3 *are linear forms in* p, q, *while* Q *is a quadratic form in* p, q, *and these forms have coefficients which are quadratic forms in the* g_{ik} *and their first order partial derivatives.*

Proof of (ii). If (1) is of class C^2 and if the partial differentiations, occurring in (3), are carried out, it is seen that the formula (3) for K can be put into the form

$$Kg^4 = \tfrac{1}{2}g^2[2g_{12uv} - g_{11vv} - g_{22uu}] + b_1 g_{11} + b_2 g_{12} + b_3 g_{22},$$

where b_1, b_2, b_3 are quadratic forms in the first order partial derivatives of the g_{ik}; cf. the standard Frobenius formula for K, [2], p. 553. If, in addition, $z(u,v)$ is of class C^3, then (12) is of class C^2 and possesses, therefore, a curvature K. The function $\kappa\gamma^4$ can be calculated by replacing g_{11}, g_{12}, g_{22} in the last formula line by $\gamma_{11} = g_{11} - p^2$, $\gamma_{12} = g_{12} - pq$, $\gamma_{22} = g_{22} - q^2$, respectively. The result of this calculation turns out to be (13); cf. [1], p. 254, where this calculation is carried out in the main. The expression on the right-hand side of (13) is identical with 4 times the expression on the left-hand side of equation (4) in [1], p. 254, except that those terms of the latter equation which do not contain p, q, r, s, t have been grouped together as Kg^4 in (13) above.

It remains to consider the case in which (1) and $z(u,v)$ are of class C^1, C^2, respectively. Let (1^{m}) be the m-th element of a sequence of positive-definite line elements satisfying the conditions of (i). Let $z^m = z^m(u,v)$, where $m = 1, 2, \cdots$, be a sequence of functions of class C^3 on R, and suppose z^m and its first and second order partial derivatives tend, as $m \to \infty$, to z and its corresponding derivatives, respectively, uniformly on every compact subset of R. Let $\gamma^m{}_{ik}(u,v)$ be defined by

$$(14^{\mathrm{m}}) \qquad \gamma^m{}_{ik}du^idu^k = g^m{}_{ik}du^idu^k - (dz^m)^2.$$

Corresponding to any compact subset of R, the line element (14^{m}) is positive-definite for sufficiently large m. On those subdomains of R on which (14^{m}) is positive-definite, let κ^m be the curvature of (14^{m}), and let $d\sigma^m = \gamma^m dudv$, where $\gamma^m = \det^{\frac{1}{2}}(\gamma^m{}_{ik}) > 0$.

Let T be an admissible Jordan domain in R. Then the formula for κ^m, deduced at the beginning of this proof, and the case $\Gamma = g^4\gamma^{-3}$, $\Gamma^m = (g^m)^4(\gamma^m)^{-3}$ of (i) imply that $\lim \iint_T \kappa^m d\sigma^m$, as $m \to \infty$, exists and has the value

$$\iint_T \{-g^2(rs - t^2) + L_1r + L_2s + L_3t + Q\}\gamma^{-3}dudv + \iint_T Kg^4\gamma^{-3}dudv.$$

Hence, the last expression is the "curvatura integra" of (12); cf. (11). Consequently, the above Definition of curvature shows that (12) has a curvature and that (13) holds. This completes the proof of (ii).

3. The case $K \equiv 0$. If a positive-definite line element (1) is of class C^2 and has a curvature $K \equiv 0$ on a simply-connected domain R, then a standard result of differential geometry states that there exist functions $x = x(u, v)$, $y = y(u, v)$ of class C^2 on R such that

$$g_{ik}du^i du^k = (dx)^2 + (dy)^2; \tag{15}$$

cf., e. g., [1], pp. 221-222. In [3], p. 767, this fact was extended to the case in which C^2 is replaced by C^1 in both the assumption and assertion. However, as shown by the following theorem, much more can be proved under the mere C^1-assumption:

(iii) *If a positive-definite line element* (1) *of class* C^1 *possesses a curvature* $K \equiv 0$ *on a simply-connected domain* R, *then there exist functions* $x = x(u, v)$, $y = y(u, v)$ *of class* C^2 *on* R *satisfying* (15). *Furthermore, such functions* x, y *can be obtained by quadratures.*

The proof will follow that given in [1], pp. 221-222, for the case in which (1) is of class C^2. It will show that if (1) if of class C^n, where $n \geqq 1$, then x and y can be chosen of class C^{n+1}.

Proof of (iii). It is seen from (7) that if $K \equiv 0$, the line integral in (6) is 0 for all Jordan curves J in R which are piecewise of class C^1. Hence, there exist functions $\theta = \theta(u, v)$ of class C^1 in R satisfying

$$\theta_u = A \quad \text{and} \quad \theta_v = B. \tag{16}$$

Such a θ can be obtained by quadratures and is determined by (16) to an additive constant.

It follows from (16) and the definitions, (4) and (5), of A and B, that

$$(g_{11}^{\frac{1}{2}} \cos \theta)_v = (g_{12} g_{11}^{-\frac{1}{2}} \cos \theta + g g_{11}^{-\frac{1}{2}} \sin \theta)_u$$

and

$$(g_{11}^{\frac{1}{2}} \sin \theta)_v = (g_{12} g_{11}^{-\frac{1}{2}} \sin \theta - g g_{11}^{-\frac{1}{2}} \cos \theta)_u.$$

Hence there exist on R functions $x = x(u, v)$, $y = y(u, v)$ of class C^2 satisfying

$$x_u = g_{11}^{\frac{1}{2}} \cos \theta, x_v = g_{12} g_{11}^{-\frac{1}{2}} \cos \theta + g g_{11}^{-\frac{1}{2}} \sin \theta$$

and

$$y_u = g_{11}^{\frac{1}{2}} \sin \theta, y_v = g_{12} g_{11}^{-\frac{1}{2}} \sin \theta - g g_{11}^{-\frac{1}{2}} \cos \theta,$$

and these functions can be obtained by quadratures. The relation (15), to be proved, is an immediate consequence of the last two formula lines. This proves (iii).

In the proofs of (I) and (II), theorem (iii) will allow the avoidance of the conformal normal form for (12), used in [2], pp. 555, 561, 562. This will permit the omission of the Hölder condition in (II).

4. *Proof of* (I). Following a device of Weingarten ([7], pp. 3-4; cf. [1], pp. 253-254), let a function $z = z(u, v)$ be sought which is of class C^2 in a vicinity of $(u, v) = (0, 0)$ and makes (12) a positive-definite line element with curvature $K \equiv 0$. In view of (ii), this leads to a Monge-Ampère partial differential equation, obtained by equating to 0 the right-hand side of (13).

The smoothness assumptions of (I) show that in this equation the coefficients, which are known functions of (u, v), are of class $C^0(\lambda)$. The assumption $K > 0$ means that the Monge-Ampère partial differential equation is of elliptic type; cf. the assumptions of the Theorems, p. 554 and the lemmas on pp. 556-557 of [2]. Hence, Lemma 2 in [2], p. 557, supplies the existence of solutions $z = z(u, v)$ of class $C^2(\mu)$ in a sufficiently small neighborhood of $(u, v) = (0, 0)$. The remarks on pp. 560-561 in [2] show that solutions z can be so chosen that (12) becomes positive-definite in a vicinity of $(0, 0)$.

If $z = z(u, v)$ denotes such a solution, then (iii) states that there exist functions $x = x(u, v)$, $y = y(u, v)$ of class C^2 in a vicinity of $(0, 0)$ satisfying $\gamma_{ik} du^i du^k = (dx)^2 + (dy)^2$. Actually, the proof of (iii) shows that x, y can be chosen of class $C^2(\mu)$, since the γ_{ik} are of class $C^1(\mu)$. In view of (12), the functions x, y, z form a solution of the system of partial differential equations (2). This completes the proof of (I).

5. *Proof of* (II) *and* (III). It is clear from the preceding proof that, in order to prove (II), it is sufficient to verify, under the conditions of (II), the following fact: The Monge-Ampère equation obtained by equating to 0 the right-hand side of (13), has, in a vicinity of $(u, v) = (0, 0)$, solutions $z = z(x, y)$ of class C^2 which make (12) positive-definite. The smoothness assumptions of (II) mean that the known coefficient functions of (u, v) in this Monge-Ampère equation are continuous near $(u, v) = (0, 0)$, while the hypothesis $K < 0$ implies that this equation is of hyperbolic type.

The existence statement supplied by Lemma 2 in [2] is not directly applicable. However, it is easy to see that, *in the hyperbolic case*, the proof of Lemma 2 (and that of Lemma 1, on which Lemma 2 depends) remains valid if the Hölder conditions occurring there are ignored. In fact, in the proofs on pp. 556-560 in [2], it is only necessary to interpret the symbol $| f |_\mu$ as the maximum of $| f |$ on the circle $u^2 + v^2 \leqq r^2$ and, as before, $\| w \|$ as greatest of the six numbers $| w |_\mu$, $| w_u |_\mu$, $| w_v |_\mu$, $| w_{uu} |_\mu$, $| w_{uv} |_\mu$, $| w_{vv} |_\mu$. This proves (II).

The remark following Lemma 2 in [2] remains valid if the Hölder conditions there are ignored. This means that if the coefficient functions in the hyperbolic Monge-Ampère equation are of class C^{n-1} (as under the assumptions of (III)), then the solutions $z = z(u, v)$ furnished by the above existence proof are of class C^{n+1}. In addition, the remark following the statement of (iii), above, shows that the corresponding functions $x = x(u, v)$ and $y = y(u, v)$ are of class C^{n+1}. This proves (III).

Paris, France and The Johns Hopkins University.

REFERENCES.

[1] G. Darboux, *Leçons sur la théorie générale des surfaces*, vol. 3, Paris, 1894.

[2] P. Hartman and A. Wintner, "On the embedding problem in differential geometry," *American Journal of Mathematics*, vol. 72 (1950), pp. 553-563.

[3] ——— and A. Wintner, "On the fundamental equations of differential geometry," *ibid.*, vol. 72 (1950), pp. 757-774.

[4] ——— and A. Wintner, "On the problems of geodesics in the small," *ibid.*, vol. 73 (1951), pp. 132-148.

[5] ——— and A. Wintner, "On the asymptotic curves of a surface," *ibid.*, vol. 73 (1951), pp. 149-172.

[6] E. R. van Kampen, "The theorems of Gauss-Bonnet and Stokes," *ibid.*, vol. 60 (1938), pp. 129-138.

[7] J. Weingarten, "Ueber die Theorie der auf einander abwickelbaren Oberflächen," *Festschrift der Technischen Hochschule zu Berlin*, 1884.

[8] H. Weyl, "Ueber die Bestimmung einer geschlossenen konvexen Fläche durch ihr Linienelement," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 61 (1915), pp. 40-72.

ON AN OSCILLATION CRITERION OF LIAPOUNOFF.*

By PHILIP HARTMAN ** and AUREL WINTNER

1. In the differential equation

$$x'' + q(t)x = 0, \tag{1}$$

let $q(t)$ be a real-valued, continuous function on a closed interval, say on $[0, T]$. Consider only the real-valued, non-trivial ($\not\equiv 0$) solutions $x(t)$ of (1). If no such solution has more than one zero on $[0, T]$, let (1) be called disconjugate on $[0, T]$. A general approach to disconjugate equations (1) was considered in [3]; it easily leads to various criteria of explicit nature, in particular to the following criterion, which goes back to Liapounoff (cf. [1]): If

$$T\int_0^T |q(t)|\,dt \leqq 4, \tag{2}$$

then (1) is disconjugate on $[0, T]$. It is also known ([2]; cf. also [1]) that the 4 in (2) cannot be improved to any $4 + \epsilon > 4$ (if $q(t)$ is unspecified). But it will turn out below that (2) can greatly be improved in another direction.

First, if the factor T in (2) is written beneath the integral sign and is then diminished to t, where $0 \leqq t \leqq T$, it is natural to ask whether there exists a positive absolute constant having the property that (1) must be disconjugate on $[0, T]$ whenever the value of

$$\int_0^T t\,|q(t)|\,dt \tag{3}$$

does not exceed that absolute constant. It turns out that such an absolute constant exists, and that its best value is 1. In other words, (1) must or need not be disconjugate on $[0, T]$ according as the value of (3) does not or does exceed 1.

---

* Received December 4, 1950.

** John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

It will also follow that if (3) is replaced by

$$\int_0^T t^a \, | \, q(t) \, | \, dt, \tag{4}$$

then there does not exist for any $T > 0$ and for any $a > 1$ a positive $C = C(a, T)$ having the property that (1) must be disconjugate on $[0, T]$ whenever the value of (4) is less than $C(a, T)$.

2. Clearly, neither is the sufficiency of

$$\int_0^T t \cdot | \, q(t) \, | \, dt \leqq 1 \tag{5}$$

contained in that of (2), nor is the converse true. But the sufficiency of both (2) and (5) is contained in part (i), and the final nature of the respective absolute constants, 4 and 1, in part (ii), of the following theorem:

Let $r = r(t)$ *be a positive, continuous function on the open interval* $(0, T)$.

(i) *Whenever* $q(t)$ *is a continuous function on the closed interval* $[0, T]$ *satisfying*

$$\int_{+0}^{T-0} r(t) \, | \, q(t) \, | \, dt \leqq T \operatorname*{g.l.b.}_{0<t<T} \{ r(t)/t(T-t) \}, \tag{6}$$

the equation (1) *must be disconjugate on* $[0, T]$, *but*

(ii) *the assertion of* (i) *becomes false for every* $r(t)$ (*and for every* T) *if the bound on the right of* (6) *is increased by* ϵ, *where* $\epsilon > 0$ *is arbitrary.*

Clearly, (2) and (5) result by choosing $r(t) = 1$ and $r(t) = t$, respectively. More generally, let $r(t) = t^a$, where $a \geqq 0$. Then the integral on the left of (6) becomes (4), while the bound on the right of (6) has the value 0, 1 or $T^{a-1}(2-a)^{2-a}/(1-a)^{1-a}$ according as $a > 1$, $a = 1$ or $a < 1$. In the more symmetric case, where $r(t) = t^a$ is replaced by $r(t) = t^a(T-t)^a$, $0 \leqq t \leqq T$, the bound on the right of (6) has the value 0 or $4^{1-a}T^{2a-1}$ according $a > 1$ or $a \leqq 1$. Thus the assertion made after (4), according to which there cannot exist a $C = C(a, T) > 0$ if $a > 1$, is obvious.

3. Let (6^+) denote the inequality which results if $| \, q(t) \, |$ in (6) is replaced by the (non-negative, continuous) function $q^+(t) = \max(0, q(t))$. Then (6) implies (6^+), but not conversely. The assertion of (i) remains, however, true if its assumption (6) is relaxed to (6^+). But this refinement of (i) is contained in (i) itself. In order to see this, it is sufficient to apply (i), as it stands, to the differential equation $x'' + q^+(t)x = 0$. Then if (6^+) holds, this differential equation is disconjugate on $[0, T]$. Hence (1) is also, since $q(t) \leqq q^+(t)$. In fact, if (1_1), (1_2) denote the differential equations (1) which belong to coefficient functions $q = q_1$, $q = q_2$ satisfying $q_1(t) \leqq q_2(t)$, then it is clear from Sturm's comparison theorem that (1_1) must be disconjugate on $[0, T]$ if (1_2) is.

This Sturmian argument also shows that it is sufficient to prove (i) for non-negative $q(t) = |q(t)|$. In addition, it can be assumed that $T = 1$ (as seen if t in (1) and (6) is replaced by t/T). But if $T = 1$ and $q(t) \geqq 0$, then (6) reduces to

$$\int_{+0}^{1-0} r(t)q(t)\,dt \leqq \gamma_r, \tag{8}$$

where

$$\gamma_r = \underset{0<t<1}{\text{g. l. b.}} \{r(t)/t(1-t)\} \tag{9}$$

and, by assumption,

$$q(t) \geqq 0 \text{ on } [0,1] \text{ and } r(t) > 0 \text{ on } (0,1). \tag{10}$$

Thus it is sufficient to prove (i)-(ii) for (1) under the normalizations (8)-(10).

The proof of (i) will lead to a generalization of an inequality of Beurling (cf. [1]), while (ii) will be proved by an adaptation of a method of Borg [1]. In principle, both (i) and (ii) must be deducible from the general criterion used in [3].

4. *Proof of* (i). If $\gamma_r = 0$, then (8) and (10) imply $q(t) \equiv 0$, and so (1) is disconjugate. Thus, it remains to consider the case $\gamma_r > 0$. It is clear from (8), (9) and (10) that

$$\int_{+0}^{1-0} r(t)q(t)\,dt \geqq \gamma_r \int_0^1 t(1-t)q(t)\,dt. \tag{11}$$

Hence, since $\gamma_r > 0$, it is sufficient to prove (i) for the case $r(t) = p(t)$, where $p(t) = t(1-t)$. In fact, (9) shows that $\gamma_p = 1$. Accordingly, (i) will be proved if it is shown that if $q(t) \geqq 0$, then

$$\int_0^1 t(1-t)q(t)\,dt > 1 \tag{12}$$

must hold whenever (1) possesses some solution $x(t) \not\equiv 0$ which has at least two zeros on $[0,1]$.

Suppose that there exists such a solution $x(t)$, and let $t=a$ and $t=b$, where

$$0 \leqq a < b \leqq 1, \tag{13}$$

be two zeros of it. Since its zeros cannot cluster on $[0,1]$, and since $x(t)$ can be replaced by $-x(t)$, it can be supposed that $x(t)$ is positive on the open interval (a,b). Then, by (1),

$$\int_a^b t(1-t)q(t)\,dt = \int_{a+0}^{b-0} -t(1-t)x''(t)/x(t)\,dt. \tag{14}$$

On the other hand, since $q(t) \geqq 0$, it is seen from (1) that the graph of $x=x(t)$ is a convex arch over (a,b). Since $x(a)=0$ and $x(b)=0$ (hence, $x(t)\equiv 0$ being excluded, $x'(a)\neq 0$ and $x'(b)\neq 0$), this convexity implies that

$$t/x(t) \geqq 1/x'(a) > 0 \text{ and } (1-t)/x(t) \geqq -1/x'(b) > 0, \tag{15}$$

if t is on the interval (a,b). Finally, if c is any point of this interval, then it is seen from (13) that $1-t > 1-c$ or $t > c$ according as t is on (a,c) or on (c,b). It follows therefore from (15) that the integral on the right of (14) is greater than (not equal to)

$$-(1-c)\int_a^c x''(t)\,dt/x'(a) + c\int_c^b x''(t)\,dt/x'(b). \tag{16}$$

Hence, this lower estimate holds for the integral on the left of (14).

This lower estimate is valid for every choice of c on (a,b). Let c now be so chosen that $x'(c)=0$ (such a choice is possible, since $x(a)=0$ and $x(b)=0$). Then (16) reduces to

$$-(1-c)(0-x'(a))/x'(a) + c(x'(b)-0)/x'(b),$$

or simply to $(1-c)+c=1$. Hence, the lower estimate, mentioned after (16), becomes

$$\int_a^b t(1-t)q(t)\,dt > 1. \tag{17}$$

Since $q(t) \geqq 0$, it is seen from (13) that (17) implies (12).

This completes the proof of (i).

5. *Proof of* (ii). Assume the normalizations (8)-(10). Let s and ϵ be arbitrary numbers satisfying $0 < \epsilon < s < 1-\epsilon$, and let $x(t)$, where $0 \leqq t \leqq 1$, be any function having the following properties: $x(t)$ has a continuous, non-positive second derivative on $[0, 1]$, is positive on $(0, 1)$, finally

$$x(t) = t \text{ on } [0, s-\epsilon] \text{ and } x(t) = (1-t)(s-\epsilon)/(1-s-\epsilon) \text{ on } [s+\epsilon, 1].$$

In terms of this $x(t) = x(t; s, \epsilon)$, define a continuous $q(t) = q(t; s, \epsilon)$ by (1), by placing $q = -x''/x$. Then $q(t)$ is non-negative on $[0, 1]$ and vanishes identically on both sub-intervals $[0, s-\epsilon]$, $[s+\epsilon, 1]$ of $[0, 1]$. Since, by the last formula line, the solution $x(t)$ of (1) has two zeros on $[0, 1]$ (but does not vanish identically), the assertion of (ii) will be proved if it is shown that the value of the integral on the left of (8) can be made less than any number which exceeds the value (9), if s and ϵ are suitably chosen.

The last formula line and (1) show that the integral on the left of (8) is identical with

$$\int_{s-\epsilon}^{s+\epsilon} -r(t)x''/x(t)\,dt. \tag{18}$$

But the value of (18) is less than $1/(s-\epsilon)$ times the value of the integral of $-r(t)x''(t)$ over the interval $[s-\epsilon, s+\epsilon]$, since $x(t) \geqq s-\epsilon$ on this interval. Hence, if $m = m(s, \epsilon)$ denotes the maximum of $r(t)$ on $[s-\epsilon, s+\epsilon]$, then the value of (18) is less than $m/(s-\epsilon)$ times

$$\int_{s-\epsilon}^{s+\epsilon} -x''(t)\,dt = x'(s-\epsilon) - x'(s+\epsilon).$$

But the last difference is identical with $1 + (s-\epsilon)/(1-s-\epsilon)$, as seen from the definition of $x(t)$. Consequently, the value of (18) is less than that of the ratio

$$m(1-2\epsilon)/\{(s-\epsilon)(1-s-\epsilon)\}, \text{ where } m = m(s, \epsilon). \tag{19}$$

Since $r(t)$ is continuous at $t = s$, it is clear from the definition of m that the ratio (19) can be brought arbitrarily close to $r(s)/\{s(1-s)\}$ by choosing ϵ small enough. Since s is an arbitrary point on $(0, 1)$, it follows that, if ϵ and s are suitably chosen, the value of (19) will come arbitrarily close to the constant (9). Consequently, the value of (18) can be made less than any number that is greater than the constant (9). This proves (ii).

PARIS, FRANCE AND THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] G. Borg, "On a Liapounoff criterion of stability," *American Journal of Mathematics*, vol. 71 (1949), pp. 67-70.

[2] E. R. van Kampen and A. Wintner, "On an absolute constant in the theory of variational stability," *ibid.*, vol. 59 (1937), pp. 270-274.

[3] A. Wintner, "On the non-existence of conjugate points," *ibid.*, vol. 73 (1951), pp. 368-380.

BOOLEAN ALGEBRAS WITH OPERATORS.*

PART I.

By Bjarni Jónsson and Alfred Tarski.

Introduction.

In the last few years various new kinds of algebraic systems have been defined and studied which can be referred to collectively as *Boolean algebras with operators.* These new algebraic systems have been obtained by enriching Boolean algebras by means of some new operations which are assumed to be additive, i. e., distributive under ordinary Boolean addition. As examples closure algebras, projective algebras, and relation algebras [1] may be mentioned. Like most other algebraic systems which have been studied in modern algebra, all these systems have arisen as results of generalizations of specific models studied in various parts of mathematics. Whenever a new kind of algebraic system originates in this way, the problem presents itself whether the abstract characterization is adequate in the sense that every system under discussion is isomorphic to one of the original models; this is the so-called representation problem for the given class of algebraic systems. As is well known, this problem has been completely solved for Boolean algebras without operators by showing that every Boolean algebra is isomorphic to an algebra formed by a field of sets under the set-theoretical operations of addition (formation of unions) and multiplication (formation of intersections).[2] It is also well known that this representation theorem is a simple consequence of (and trivially equivalent to) the so-called extension theorem by which every Boolean algebra can be extended to a complete and atomistic Boolean algebra. The aim of this work is to extend these results to arbitrary Boolean algebras with operators. We first establish the *extension theorem* by showing that every

---

* Received March 13, 1950.

[1] Axiomatic studies of closure algebras, projective algebras, and relation algebras were first undertaken in McKinsey-Tarski [1], Everett-Ulam [1], and Tarski [2], respectively. The operator lattices discussed in Duffin-Pate [1] constitute still another example of Boolean algebras with operators. The general notion of a Boolean algebra with operators was first introduced and the main result of the present work were briefly stated in Jónsson-Tarski [1] (the abstract of a paper read at a meeting of the American Mathematical Society in November 1947). The numbers in brackets refer to the bibliography which follows the introduction.

[2] Cf. Stone [2], Theorem 67.

algebra of this kind can be extended to a complete and atomistic Boolean algebra where, moreover, all the operations are completely additive, i. e., distributive under addition of infinitely many elements. We also show that every equation involving Boolean addition and multiplication and the operators, which is identically satisfied in the original algebra, is also identically satisfied in the extended algebra. Hence, if the original algebra belongs to a class of algebras characterized by means of postulates all of which are algebraic identities (as, e. g., in the case of closure algebras), then the extended algebra belongs to the same class. This extension theorem implies a general *representation theorem* for Boolean algebras with operators; roughly speaking every such algebra is isomorphic to an algebra formed by a field of sets with the usual set-theoretical operations, and with operators defined as images under certain relations between elements of the universal set (the largest set) of this field of sets, the notion of the image under a relation being a generalization of the notion of the image under a function. We can also express this by saying that every Boolean algebra with operators is isomorphic to a subalgebra of the complex algebra of an algebraic system formed by a set and certain relations between its elements. When applying these general results to special Boolean algebras with operators, e. g., to those mentioned at the beginning of this introduction, we do not always obtain the natural representation theorem suggested by the origin of these algebras. In some cases a natural representation theorem can be obtained by combining our results with some arguments known from the literature.

After deriving at the beginning of Section 1 certain rather elementary theorems concerning additive functions, we define the notion of conjugate functions and investigate its basic properties. In the last part of this section we introduce the notion of a perfect extension of a Boolean algebra and restate the extension theorem for Bolean algebras without operators in terms of this concept.

In Section 2 the extension theorem for Boolean algebras with operators is proved. This result is used in Section 3 to obtain the representation theorem for Boolean algebras with operators. In the last part of this section both theorems are applied to closure algebras and to cylindric algebras (which present a generalization of the projective algebras mentioned above).

In Section 4 the extension theorem is applied to relation algebras. From this we obtain a new kind of representation theorem for relation algebras which is not a special case of our general representation theorem and which in certain rather special cases proves to supply a natural representation for these algebras.

In the fifth and last section we study the complex algebras of Brandt groupoids and groups. Here we also introduce the notion of a generalized Brandt groupoid and show that a relation algebra has a natural representation if, and only if, it is isomorphic to a subalgebra of the complex algebra of a generalized Brandt groupoid.

BIBLIOGRAPHY.

G. Birkhoff [1]. "Rings of sets," *Duke Mathematical Journal*, vol. 3 (1937), pp. 443-454.

G. Birkhoff [2]. "Subdirect unions in universal algebra," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 764-768.

H. Brandt [1]. "Über eine Verallgemeinerung des Gruppenbegriffes," *Mathematische Annalen*, vol. 96 (1926-1927), pp. 360-366.

J. R. Büchi [1]. "Die Boole'sche Partialordnung und die Paarung von Gefuegen," *Portugaliae Mathematica*, vol. 7 (1948), pp. 119-180.

L. H. Chin and A. Tarski [1]. *Distributive and modular laws in relation algebras.*

L. H. Chin and A. Tarski [2]. "Remarks on projective algebras," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 80-81.

R. J. Duffin and R. S. Pate [1]. "An abstract theory of Jordan-Hölder composition series," *Duke Mathematical Journal*, vol. 10 (1943), pp. 743-750.

C. J. Everett and S. Ulam [1]. "Projective algebras I," *American Journal of Mathematics*, vol. 68 (1946), pp. 77-88.

B. Jónsson and A. Tarski [1]. "Boolean algebras with operators," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 79-80.

B. Jónsson and A. Tarski [2]. "Representation problems for relation algebras," *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 80 and 1192.

R. C. Lyndon [1]. "The representation of relation algebras," *Annals of Mathematics*, vol. 51 (1950), pp. 707-729.

J. C. C. McKinsey [1]. "On the representation of projective algebras," *American Journal of Mathematics*, vol. 70 (1948), pp. 375-384.

J. C. C. McKinsey [2]. "Postulates for the calculus of binary relations," *Journal of Symbolic Logic*, vol. 5 (1940), pp. 85-97.

J. C. C. McKinsey and A. Tarski [1]. "The algebra of topology," *Annals of Mathematics*, vol. 45 (1944), pp. 141-191.

R. Sikorski [1]. "A theorem on extension of homomorphisms," *Annales de la Société Polonaise de Mathématique*, vol. 21 (1949), pp. 332-335.

R. Sikorski [2]. "On the inducing of homomorphisms by mappings," *Fundamenta Mathematicae*, vol. 36 (1949), pp. 7-22.

M. H. Stone [1]. "Applications of the theory of Boolean rings to general topology," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 375-481.

M. H. Stone [2]. "The theory of representations for Boolean algebras," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 37-111.

A. Tarski [1]. "A fixpoint theorem for lattices and its applications," *Bulletin of the American Mathematical Society*, vol. 55 (1949), pp. 1051-1052.

A. Tarski [2]. "On the calculus of relations," *Journal of Symbolic Logic*, vol. 6 (1941), pp. 73-89.

A. Tarski [3]. "Sur quelques propriétés caractéristiques des images d'ensembles," *Annales de la Société Polonaise de Mathématique*, vol. 6 (1927), pp. 127-128.

Notations.

Throughout this work the letters j, m, n, p, q, and r will denote non-negative integers. By a sequence we shall, unless otherwise specified, mean a finite sequence. If x is an m-termed sequence, then for $j < m$ we understand x_j to be the $(j+1)$-st term of x. If x is an m-termed sequence and y is an n-termed sequence, then $x^\frown y$ denotes the $(m+n)$-termed sequence z such that $z_j = x_j$ for $j < m$ and $z_j = y_{j-m}$ for $m \leq j < m+n$. No distinction is made between an element a and the one-termed sequence x such that $x_0 = a$. Thus, given two elements a and b, we write alternatively $a^\frown b$ or $\langle a, b\rangle$ for the two-termed sequence, or ordered couple, x such that $x_0 = a$ and $x_1 = b$. If X is an m-termed sequence of sets, then $X_0 \times X_1 \times \cdots \times X_{m-1}$ denotes the set of all m-termed sequences x such that $x_j \varepsilon X_j$ (x_j is in X_j) for $j < m$; in case $X_j = A$ for every $j < m$, we agree that $A^m = X_0 \times X_1 \times \cdots \times X_{m-1}$.

A family A of sets is called a set-field if the union $X \cup Y$ and the difference $X - Y$ of any two sets X and Y in A are themselves members of A and if there exists a largest set in A, i. e., a set $U \varepsilon \mathsf{A}$ such that $X \subseteq U$ whenever $X \varepsilon \mathsf{A}$. This set U can be defined as the union of all sets in A:

$$U = \bigcup_{X \varepsilon \mathsf{A}} X;$$

it is called the universal set of A. From this definition it follows that the intersection $X \cap Y$ of any two members of A is itself a member of A and that the empty set Λ belongs to A.

As is well known, m-ary relations can be thought of as sets of m-termed sequences. We can therefore apply to them the usual set-theoretical operations. In the case of binary relations we shall need two other operations, namely relative multiplication and conversion. The *relative product* of two binary relations R and S, denoted by $R|S$, is the set of all ordered couples $\langle x, y\rangle$ such that $\langle x, z\rangle \varepsilon R$ and $\langle z, y\rangle \varepsilon S$ for some element z, while R^{-1}, the *converse* of R, is the set of all ordered couples $\langle x, y\rangle$ such that $\langle y, x\rangle \varepsilon R$. A function of rank m (or a function in m variables) may be thought of as a special kind of $(m+1)$-ary relation. Thus an $(m+1)$-ary relation R is a function if, and only if, for each m-termed sequence x there exists at most one element a such that $x^\frown a \varepsilon R$.

If f is a function and A is a subset of the domain of f, then f/A will denote the function f with its domain restricted to A, i. e., f/A is the function g with domain A such that $g(x) = f(x)$ whenever $x \varepsilon A$. Given three sets A, B, and C, a function f on B^m to C, and an m-termed sequence g of functions on A^n to B, we shall understand by $f[g]$ or $f[g_0, g_1, \cdots, g_{m-1}]$ the function h on A^n to C such that

$$h(x) = f(g_0(x), g_1(x), \cdots, g_{m-1}(x)) \quad \text{whenever } x \varepsilon A^n.$$

Under the same conditions, $f\langle g\rangle$ or $f\langle g_0, g_1, \cdots, g_{m-1}\rangle$ will denote the function k on A^{mn} to C such that

$$k(x^{(0)\frown}x^{(1)\frown}\cdots{}^\frown x^{(m-1)}) = f(g_0(x^{(0)}), g_1(x^{(1)}), \cdots, g_{m-1}(x^{(m-1)}))$$
$$\text{whenever } x^{(0)}, x^{(1)}, \cdots, x^{(m-1)} \varepsilon A^n.$$

We shall refer to $f[g]$ and $f\langle g\rangle$ as the *composition* and the *superposition*, respectively, of f and g. The notion of superposition could clearly be extended to functions $g_0, g_1, \cdots$ of different ranks, but this generalization is not important for our purposes. A function f of rank m will be called an *identity function* if, for some $j < m$, we have $f(x) = x_j$ whenever x is an m-termed sequence belonging to the domain of f.

By an *algebra* we shall mean a finite, infinite, or transfinite sequence

$$\mathfrak{A} = \langle A, f_0, f_1, \cdots, f_\xi, \cdots\rangle$$

where A is a non-empty set and each f_ξ is a function of some finite rank m_ξ with domain A^{m_ξ} and counterdomain a subset of A. In this context the term 'operation' is sometimes used instead of the term 'function'; $f_0, f_1, \cdots, f_\xi, \cdots$ are referred as *fundamental operations* of $\mathfrak{A}$. By a subalgebra of $\mathfrak{A}$ we understand a sequence

$$\mathfrak{B} = \langle B, f_0/B^{m_0}, f_1/B^{m_1}, \cdots, f_\xi/B^{m_\xi}, \cdots\rangle$$

where B is a non-empty subset of A such that $f_\xi(x) \varepsilon B$ whenever $x \varepsilon B^{m_\xi}$. Sometimes we shall, however, speak of the subalgebra

$$\langle B, f_0, f_1, \cdots, f_\xi, \cdots\rangle$$

of $\mathfrak{A}$, meaning actually the algebra $\mathfrak{B}$ described above. Frequently we consider an algebra as a sequence where the first term is a non-empty set A

while the remaining terms are functions or certain distinguished elements of A. This is justified since we could replace each such element a by a function f of rank 0 such that $f(\Lambda) = a$. Thus we shall consider a Boolean algebra as a system

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1\rangle$$

where $0, 1 \varepsilon A$, and $+$ and $\cdot$ are binary operations subject to certain familiar conditions.

We shall employ such familiar notions from general algebra as those of isomorphism ($\mathfrak{A} \cong \mathfrak{B}$), homomorphism, cardinal—or direct—product of two algebras ($\mathfrak{A} \times \mathfrak{B}$), or of arbitrarily many algebras, cardinal—or direct—power, simple algebra, and indecomposable—or directly indecomposable—algebra. Also the less familiar notions of a similarity class of algebras, an algebraic function over a similarity class, and an equationally definable class of algebras will play some part in our discussion.[3] Two algebras

$$\mathfrak{A} = \langle A, f_0, f_1, \cdots, f_\xi, \cdots\rangle \text{ with } \xi < \alpha$$

and

$$\mathfrak{B} = \langle B, g_0, g_1, \cdots, g_\xi, \cdots\rangle \text{ with } \xi < \beta$$

(where α and β are any finite or transfinite ordinals) are called *similar* if $\alpha = \beta$, and f_ξ and g_ξ are functions of the same rank for every $\xi < \alpha$. The class consisting of all algebras which are similar to a certain given algebra is referred to as a *similarity class*. Given a similarity class K, a function f is called *algebraic over* K if roughly speaking, f correlates with every algebra $\mathfrak{A}$ in K a function $f_{\mathfrak{A}}$ on A^m to A (m being some natural number) which is obtained by iterated composition from the fundamental operations of $\mathfrak{A}$. If, e. g., K is the similarity class to which the Boolean algebras belong, and if, for every algebra $\mathfrak{A} = \langle A, +, 0, \cdot, 1\rangle$ in K, $f_{\mathfrak{A}}$ is a function on A^3 to A such that

$$f_{\mathfrak{A}}(x, y, z) = (x + y) \cdot z \text{ for any } x, y, z \varepsilon A,$$

then f is an algebraic function over K. A subclass L of K is called *equationally definable* if it can be characterized by means of a finite or infinite system of axioms all of which have the form of equations between algebraic functions over K. Speaking more precisely, L is equationally definable if there is a set S of couples $\langle f, g\rangle$ of algebraic functions over K such that an algebra $\mathfrak{A}$ in K belongs to L if, and only if, $f_{\mathfrak{A}} = g_{\mathfrak{A}}$ for every couple $\langle f, g\rangle \varepsilon S$.

---

[3] For these notions see McKinsey-Tarski [1] where further bibliographic references (in particular to papers of G. Birkhoff) can also be found.

Thus, for instance, Boolean algebras as they are conceived here do not constitute an equationally definable class; if, however, we agree to treat Boolean algebras as systems

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, ^{-} \rangle$$

with an additional unary operation $^{-}$ (the operation of complementation), then the class of Boolean algebras becomes equationally definable—though, of course, within a different similarity class.

In connection with the representation theorem we shall consider systems

$$\mathfrak{A} = \langle A, R_0, R_1, \cdots, R_\xi, \cdots \rangle$$

where A is a non-empty set and each R_ξ is a subrelation of $A^{m_\xi+1}$ for some finite m_ξ. Such systems will be called *algebras in the wider sense.*

When referring to Boolean algebras, the symbols $+$, $\cdot$, $\sum$, $\prod$, $\leq$, $\geq$, 0, 1 will have their usual meaning; the complement of an element x will be denoted by x^{-}, and the difference of two elements x and y, i. e., $x \cdot y^{-}$, by $x - y$. Even when several Boolean algebras are being considered at one time these symbols will be used for all of them, it being clear from the context which algebra is referred to each time. However, in the case of two Boolean algebras where one is a subalgebra of the other, the symbols $\sum$ and $\prod$ will always refer to the larger algebra. When applied to sequences of elements of a Boolean algebra, the symbols $+$, $\cdot$, $-$, $^{-}$, $\sum$, $\prod$, $\leq$, and $\geq$ should be interpreted in the sense of the cardinal power of the given algebra. Thus, e. g., if x and y are two such m-termed sequences, then $x \leq y$ will mean that $x_j \leq y_j$ for every $j < m$. Similarly, if f and g are functions on an arbitrary set I to a Boolean algebra, then $f \leq g$ will mean that $f(i) \leq g(i)$ for every $i \varepsilon I$.

Section 1.

Additive Functions and Conjugate Functions.

Throughout the first part of this section we consider a fixed Boolean algebra

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1 \rangle$$

and let At denote the set consisting of 0 and all the atoms of $\mathfrak{A}$.

Definition 1. 1. *A function f on A^m to A is called*

(i) *normal if, given any $j < m$ and a sequence $x \varepsilon A^m$ such that $x_j = 0$, we always have $f(x) = 0$;*

(ii) *monotonic if, given two sequences* $x, y \varepsilon A^m$ *such that* $x \leq y$, *we always have* $f(x) \leq f(y)$;

(iii) *additive if, given any* $j < m$ *and two sequences* $x, y \varepsilon A^m$ *such that* $x_p = y_p$ *whenever* $j \neq p < m$, *we always have*

$$f(x + y) = f(x) + f(y);$$

(iv) *completely additive if, given any* $j < m$, *a non-empty set* I, *and sequences* $x^{(i)} \varepsilon A^m$ *with* $i \varepsilon I$ *such that* $x_p{}^{(i)} = x_p{}^{(i')}$ *whenever* $i, i' \varepsilon I$ *and* $j \neq p < m$, *and such that* $\sum_{i \varepsilon I} x^{(i)}$ *exists, then* $\sum_{i \varepsilon I} f(x^{(i)})$ *also exists and we have*

$$f(\sum_{i \varepsilon I} x^{(i)}) = \sum_{i \varepsilon I} f(x^{(i)}).$$

Obviously, every additive function is monotonic, and every completely additive function is additive.

THEOREM 1.2. *The operations of addition and multiplication in* $\mathfrak{A}$ *are completely additive functions, the latter being also normal. Furthermore, if* f *is an identity function on* A^m *to* A, *or if* $a \varepsilon A$ *and* f *is the function on* A^m *to* A *defined by the formula*

$$f(x) = a \text{ for any } x \varepsilon A^m,$$

then f *is completely additive.*

Proof. Obvious.

THEOREM 1.3. *If* f *is an additive (or completely additive) function on* A^m *to* A *and if* g *is the function on* A^{m+n} *to* A *defined by the formula*

$$g(x^\frown y) = f(x) \text{ for any } x \varepsilon A^m \text{ and } y \varepsilon A^n,$$

then g *is additive (or completely additive).*

Proof. Obvious.

THEOREM 1.4. *If* f *is an additive (or completely additive) function on* A^{m+n} *to* A *and if* $a \varepsilon A^n$ *and* g *is the function on* A^m *to* A *defined by the formula*

$$g(x) = f(x^\frown a) \text{ for any } x \varepsilon A^m,$$

then g *is additive (or completely additive).*

Proof. Obvious.

THEOREM 1.5. *A function* f *on* A *to* A *is additive (or completely* *additive) if, and only if, there exist a normal and additive (or completely additive) function g on A to A and an element $a \varepsilon A$ such that*

$$f(x) = g(x) + a \text{ for any } x \varepsilon A.$$

Similarly, a function f on A^2 to A is additive (or completely additive) if, and only if, there exist a normal and additive (or completely additive) function g on A^2 to A, two normal and additive (or completely additive) function h and k on A to A, and an element $a \varepsilon A$ such that

$$f(x) = g(x) + h(x_0) + k(x_1) + a \text{ for any } x \varepsilon A^2.$$

Proof. Assume that f is an additive function on A to A. We let

$$a = f(0), \text{ and } g(x) = f(x) - a \text{ for every } x \varepsilon A.$$

By 1. 1, g is normal and additive, and we have

$$f(x) = g(x) + a \text{ for every } x \varepsilon A. \tag{1}$$

(Similarly, if f is completely additive, then g is also completely additive.) Conversely, if $a \varepsilon A$, g is a normal and additive (or completely additive) function on A to A, and (1) holds, then f is additive (or completely additive) by 1. 1.

To prove the second part of the theorem we proceed in an analogous manner; we let

$$a = f(0, 0),$$
$$h(x) = f(x, 0) - a \text{ and } k(x) = f(0, x) - a \text{ for every } x \varepsilon A,$$
$$g(x) = f(x) - (h(x_0) + k(x_1) + a) \text{ for every } x \varepsilon A^2.$$

Similar theorems can of course be obtained for functions of higher rank. From this we see that it would not be a serious loss of generality if we restricted our discussion to normal functions, as in fact we will do in Section 3.

THEOREM 1. 6 (GENERAL DISTRIBUTIVE LAW). *Let f be a function on A^m to A. Then the following conditions are equivalent:*

(i) *The function f is additive.*

(ii) *If $K_0, K_1, \cdots, K_{m-1}$ are finite, non-empty subsets of A and if $L = K_0 \times K_1 \times \cdots \times K_{m-1}$, then*

$$f(\sum_{x \varepsilon L} x) = \sum_{x \varepsilon L} f(x).$$

Similarly, the following conditions are equivalent:

(i′) *The function f is completely additive.*

(ii′) *If $K_0, K_1, \cdots, K_{m-1}$ are non-empty subsets of A, if $L = K$ $\times \cdots \times K_{m-1}$, and if $\sum_{x \in L} x$ exists, then $\sum_{x \in L} f(x)$ also exists and we*

$$f(\sum_{x \in L} x) = \sum_{x \in L} f(x).$$

Proof. By induction using 1.1 and 1.4.

THEOREM 1.7. *Suppose $\mathfrak{A}$ is atomistic and f is a function on A Then f is completely additive if, and only if,*

$$f(x) = \sum_{x \geq u \in At^m} f(u) \textit{ for any } x \in A^m.$$

Proof. The necessity of this condition follows from 1.6. Assu the condition is satisfied. Consider a number $j < n$, a non-empty set sequences $x^{(i)} \in A^m$ with $i \in I$ such that

$$x_p^{(i)} = x_p^{(i')} \text{ whenever } i, i' \in I \text{ and } j \neq p < m,$$

and such that $\sum_{i \in I} x^{(i)}$ exists. Let

$$L_i = \underset{u}{\mathrm{E}} [u \in At^m \text{ and } u \leq x^{(i)}] \text{ for any } i \in I, \text{ and } K = \bigcup_{i \in I} L_i.$$

Observe that

$$u \in K \text{ if, and only if, } \sum_{i \in I} x^{(i)} \geq u \in At^m.$$

Therefore

$$f(\sum_{i \in I} x^{(i)}) = \sum_{u \in K} f(u) = \sum_{i \in I} \sum_{u \in L_i} f(u) = \sum_{i \in I} f(x^{(i)}).$$

Hence f is completely additive by 1.1(iv).

THEOREM 1.8. *Suppose $\mathfrak{A}$ is atomistic. For any completely functions f and g on A^m to A we then have:*

(i) *$f \leq g$ if, and only if, $f(x) \leq g(x)$ for any $x \in A^m$.*

(ii) *$f = g$ if, and only if, $f(x) = g(x)$ for any $x \in A^m$.*

Proof. By 1.7.

THEOREM 1.9. *Let f be an additive (or completely additive) f on A^m to A and let $g_0, g_1, \cdots, g_{m-1}$ be additive (or completely ac functions on A^n to A. Then $f\langle g_0, g_1, \cdots, g_{m-1}\rangle$ is an additive (c pletely additive) function on A^{mn} to A.*

Proof. Assume that $f, g_0, g_1, \cdots, g_{m-1}$ are additive and let

$$(1) \qquad h = f\langle g_0, g_1, \cdots, g_{m-1}\rangle.$$

Suppose $j < mn$, and consider two sequences $x, y \varepsilon A^{mn}$ such that

$$(2) \qquad x_p = y_p \text{ whenever } j \neq p < mn.$$

Then there exist sequences $z^{(0)}, z^{(1)}, \cdots, z^{(m-1)}$, $u^{(0)}, u^{(1)}, \cdots, u^{(m-1)} \varepsilon A^n$ such that

$$(3) \qquad x = z^{(0)}{}^\frown z^{(1)}{}^\frown \cdots {}^\frown z^{(m-1)} \text{ and } y = u^{(0)}{}^\frown u^{(1)}{}^\frown \cdots {}^\frown u^{(m-1)}.$$

Let q be the greatest integer such that $q \leq j/n$. Then, by (2) and (3),

$$(4) \qquad z^{(r)} = u^{(r)} \text{ whenever } q \neq r < m,$$

$$(5) \qquad z_s^{(q)} = u_s^{(q)} \text{ whenever } j - nq \neq s < n.$$

Hence, by 1.1(iii),

$$(6) \qquad g_r(z^{(r)} + u^{(r)}) = g_r(z^{r}) + g_r(u^{(r)}) \text{ whenever } r < m.$$

Let the sequences $v, w \varepsilon A^m$ be defined by the formulas

$$(7) \qquad v_r = g_r(z^{(r)}) \text{ and } w_r = g_r(u^{(r)}) \text{ whenever } r < m.$$

Then, by (1) and (3),

$$h(x) = f(v) \text{ and } h(y) = f(w).$$

Therefore, by (4), (7), and 1.1(iii),

$$h(x) + h(y) = f(v + w).$$

But by (1), (3), (6), (7), and 1.1(iii) we have

$$h(x + y) = f(v + w).$$

Consequently,

$$(8) \qquad h(x + y) = h(x) + h(y).$$

Since (2) implies (8), we see by 1.1(iii) that h is additive. (In the case of complete additivity the proof is similar.)

It is not in general true that, under the hypothesis of the preceding theorem, the function $f[g_0, g_1, \cdots, g_{m-1}]$ is additive (or completely additive). In fact, it is precisely for this reason that we introduced the notion of the superposition of functions. We have, however, the following

THEOREM 1.10. *Let Φ be the smallest family of functions which satisfies the following conditions:*

(i) *If, for some m, f is an additive function on A^m to A, then $f \varepsilon \Phi$.*

(ii) *If, for some m and n, $f \varepsilon \Phi$ is a function on A^m to A and $g_0, g_1, \cdots, g_{m-1} \varepsilon \Phi$ are functions on A^n to A, then $f[g_0, g_1, \cdots, g_{m-1}] \varepsilon \Phi$.*

For a function f on A^n to A to belong to the family Φ it is necessary and sufficient that there exist, for some m, an additive function f' on A^m to A and m identity functions $h_0, h_1, \cdots, h_{m-1}$ on A^n to A such that

$$f = f'[h_0, h_1, \cdots, h_{m-1}].$$

Proof. Consider the family Φ' consisting of all functions f such that, for some m and n, f is of the form

$$f = f'[h_0, h_1, \cdots, h_{m-1}],$$

where f' is an additive function on A^m to A and $h_0, h_1, \cdots, h_{m-1}$ are identity functions on A^n to A. If we replace Φ by Φ', then (i) is clearly satisfied, and we shall show that (ii) is also satisfied, whence $\Phi \subseteq \Phi'$.

In fact, consider a function $f \varepsilon \Phi'$ on A^m to A and m functions $g_0, g_1, \cdots, g_{m-1} \varepsilon \Phi'$ on A^n to A. We then have, for some p,

$$f = f'[h_0, h_1, \cdots, h_{p-1}], \tag{1}$$

where f' is an additive function on A^p to A and $h_0, h_1, \cdots, h_{p-1}$ are identity functions on A^m to A. Let

$$k_j = h_j[g_0, g_1, \cdots, g_{m-1}] \text{ for every } j < p. \tag{2}$$

Then each k_j is a g_q for some $q < m$. Hence, for some r,

$$k_j = k'_j[l_0^{(j)}, l_1^{(j)}, \cdots, l_{r-1}^{(j)}] \text{ for } j < p \tag{3}$$

where the k'_j are additive functions on A^r to A while $l_0^{(j)}, l_1^{(j)}, \cdots, l_{r-1}^{(j)}$ are identity functions on A^n to A. The assumption that we have the same number r for all values of j is justified by 1.3. Let

$$g' = f'\langle k'_0, k'_1, \cdots, k'_{p-1}\rangle. \tag{4}$$

Then, by 1.9, g' is an additive function on A^{pn} to A. By (1) and (2) we have

$$f[g_0, g_1, \cdots, g_{m-1}] = f'[k_0, k_1, \cdots, k_{p-1}].$$

Hence, if we let

$$l = l^{(0)} \frown l^{(1)} \frown \cdots \frown l^{(p-1)},$$

then, by (3) and (4),

$$f[g_0, g_1, \cdots, g_{m-1}] = g'[l_0, l_1, \cdots, l_{pn-1}].$$

Thus $f[g_0, g_1, \cdots, g_{m-1}] \varepsilon \Phi'$. We conclude that $\Phi \subseteq \Phi'$. The inclusion in the opposite direction follows from 1. 2. Hence $\Phi = \Phi'$, and the proof is complete.

We are now about to introduce and study the notion of conjugate functions.[4] The simple definition of conjugate functions proves to be very rich in consequences. At the first glance the new notion may apepar to be rather special; it is not involved in the formulations and proofs of the main theorems of this paper. However, our general discussion will lead to conjugate functions in quite a natural manner (compare Theorem 3. 6 below); we shall also apply this notion in the discussion of cylindric algebras and relation algebras.

DEFINITION 1. 11. *Let f and g be functions on A to A. We say that g is a conjugate of f if, for any $x, y \varepsilon A$, we have*

$$f(x) \cdot y = 0 \text{ if, and only if, } g(y) \cdot x = 0.$$

If, in particular, a function f is a conjugate of itself, then we call f self-conjugate.

THEOREM 1. 12. *If f and g are functions on A to A and g is a conjugate of f, then f is a conjugate of g.*

Proof. Obvious.

In view of this theorem we shall usually say "f and g are conjugate" instead of "g is a conjugate of f."

THEOREM 1. 13. *A function f on A to A has at most one conjugate function g, and this function (if it exists) is determined by the formula*

$$g(y) = \prod_{f(x)\cdot y=0} x^- \text{ for any } y \varepsilon A.$$

Proof. Let g be a conjugate of f. By 1. 11 we have, for any elements $x, y \varepsilon A$,

$$f(x) \cdot y = 0, \text{ if, and only if, } x \leq [g(y)]^-.$$

---

[4] This notion was first discussed in Tarski [3] (though no special term to denote it was introduced there). Recently the same notion was studied (under another name) in Büchi [1], pp. 157 ff.

Thus, given an element $y \varepsilon A$, the largest element $x \varepsilon A$ for which $f(x) \cdot y = 0$ is $[g(y)]^{-}$. Consequently

$$[g(y)]^{-} = \sum_{f(x) \cdot y = 0} x \text{ for } y \varepsilon A.$$

Taking complement on both sides we obtain the desired formula for $g(y)$.

THEOREM 1. 14. *A function f on A to A has a conjugate if, and only if, the following conditions are satisfied:*

(i) *f is normal,*

(ii) *f is completely additive,*

(iii) *$\sum_{f(x) \leq y} x$ exists for every element $y \varepsilon A$.*

Proof. Suppose that f has a conjugate g. Since $g(1) \cdot 0 = 0$, we have $f(0) \cdot 1 = 0$ by 1. 11. Therefore $f(0) = 0$, and f is normal by 1. 1(i). Let I be an arbitrary non-empty set and let the elements $x_i \varepsilon A$ with $i \varepsilon I$ be such that $\sum_{i \varepsilon I} x_i$ exists. Let

$$y = \sum_{i \varepsilon I} x_i. \tag{1}$$

Consider any element $z \varepsilon A$ such that

$$f(x_i) \leq z \text{ for every } i \varepsilon I. \tag{2}$$

For each $i \varepsilon I$ we then have $f(x_i) \cdot z^{-} = 0$ and hence, by 1. 11,

$$g(z^{-}) \cdot x_i = 0 \text{ for any } i \varepsilon I.$$

By (1) this implies that $g(z^{-}) \cdot y = 0$, and therefore, by 1. 11, that

$$f(y) \leq z. \tag{3}$$

It is easy to check that every step in the derivation of (3) from (2) is reversible and, hence, that the two formulas are equivalent for any $z \varepsilon A$. Consequently

$$f(y) = \sum_{i \varepsilon I} f(x_i). \tag{4}$$

Since (1) always implies (4), f is completely additive by 1. 1(iv). An elementary transformation of the formula in 1. 13 gives

$$[g(y^{-})]^{-} = \sum_{f(x) \leq y} x \text{ for any } y \varepsilon A.$$

Therefore the right-hand member of this equation exists for every $y \varepsilon A$.

We have shown that if f has a conjugate, then conditions (i)-(iii) all hold. Conversely, assume now that (i)-(iii) are satisfied. Then the formula

$$g(y) = [\sum_{f(x)\leq y^-} x]^- \text{ for any } y \,\varepsilon\, A \tag{5}$$

defines a function g on A to A. We can write (5) in the form

$$g(y) = \prod_{f(x)\cdot y=0} x^- \text{ for any } y \,\varepsilon\, A,$$

and it follows that

$$f(x)\cdot y = 0 \text{ implies that } g(y)\cdot x = 0 \text{ for any } x, y \,\varepsilon\, A. \tag{6}$$

By (5) we have

$$[g(y)]^- = \sum_{f(x)\leq y^-} x \text{ for any } y \,\varepsilon\, A.$$

Hence by (i), (ii), and 1.1(i)(iv)

$$f([g(y)]^-) = \sum_{f(x)\leq y^-} f(x) \leq y^- \text{ for any } y \,\varepsilon\, A. \tag{7}$$

If now $x, y \,\varepsilon\, A$ are such that $g(y)\cdot x = 0$, then $x \leq [g(y)]^-$. Applying first (ii) and then (7) we obtain $f(x) \leq f([g(y)]^-)$ and $f(x) \leq y^-$. Consequently,

$$g(y)\cdot x = 0 \text{ implies that } f(x)\cdot y = 0 \text{ for any } x, y \,\varepsilon\, A. \tag{8}$$

From (6) and (8) we conclude by 1.11 that f and g are conjugate. The proof is complete.

The function f defined by the formula

$$f(x) = a\cdot x$$

is of course selfconjugate. Hence Theorem 1.14 implies as a particular case the familiar distributive law for binary multiplication under general addition (for $\cdot$ under $\sum$).

THEOREM 1.15. *If f and g are functions on A to A, then the following conditions are equivalent:*

(i) *f and g are conjugate.*

(ii) *$f(x - g(y)) \leq f(x) - y$ and $g(y - f(x)) \leq g(y) - x$ for any $x, y \,\varepsilon\, A$.*

(iii) *f and g are normal, and we have $f(x)\cdot y \leq f(x\cdot g(y))$ and $g(y)\cdot x \leq g(y\cdot f(x))$ for any $x, y \,\varepsilon\, A$.*

Proof. Suppose f and g are conjugate, and let $x, y \,\varepsilon\, A$. Then $f(x - g(y)) \leq f(x)$ by 1.1(iv) and 1.14, while by 1.11 we have $f(x - g(y))\cdot y = 0$. Hence

$$f(x - g(y)) \leq f(x) - y. \tag{1}$$

Analogously we have

$$(2) \qquad g(y-f(x)) \leq g(y)-x.$$

Assume, conversely, that (1) and (2) hold for any elements $x, y \varepsilon A$. If $f(x)\cdot y=0$, then $y=y-f(x)$ and we have, by (2),

$$g(y)=g(y-f(x)) \leq g(y)-x \leq x^{-},$$

so that $g(y)\cdot x=0$. Similarly, by (1), $g(y)\cdot x=0$ implies that $f(x)\cdot y=0$. Hence f and g are conjugate functions. Thus (i) and (ii) are equivalent.

Suppose f and g are conjugate functions. Then f and g are normal by 1. 14. By 1. 1(iv) and 1. 14 we also have, for any elements $x, y \varepsilon A$,

$$f(x)=f(x-g(y))+f(x\cdot g(y)).$$

Since (i) implies (ii), we infer that

$$f(x) \leq (f(x)-y)+f(x\cdot g(y)).$$

Consequently

$$f(x)\cdot y \leq (f(x)-y)\cdot y+f(x\cdot g(y))\cdot y \leq f(x\cdot g(y)).$$

Similarly,

$$g(y)\cdot x \leq g(y\cdot f(x)).$$

Conversely, assume that (iii) holds. If $x, y \varepsilon A$ are such that $f(x)\cdot y=0$, then

$$g(y)\cdot x \leq g(y\cdot f(x))=g(0)=0$$

Similarly, $g(y)\cdot x=0$ implies that $f(x)\cdot y=0$. Hence f and g are conjugate by 1. 11 and the proof is complete.

The next three theorems are of a rather special nature, but they will have added significance in connection with Theorems 3. 5 and 3. 7.

THEOREM 1. 16. *If f is a function on A to A, then the following conditions are equivalent:*

(i) *f is selfconjugate.*

(ii) *f is additive, and $f([f(x)]^{-})\cdot x=0$ for any $x \varepsilon A$.*

(iii) *f is normal, and $f(x)\cdot y \leq f(x\cdot f(y))$ for any $x, y \varepsilon A$.*

Proof. If (ii) is satisfied, then by 1. 1(iii).

$$f(0) \leq f([f(1)]^{-})\cdot 1=0.$$

Hence $f(0)=0$, and f is normal by 1. 1(i). Furthermore, if $x, y \varepsilon A$, then 1. 1(iii) gives with the aid of (ii)

$$f(x)\cdot y = f(x\cdot f(y))\cdot y + f(x\cdot[f(y)]^{-})\cdot y$$
$$\leq f(x\cdot f(y)) + f([f(y)]^{-})\cdot y = f(x\cdot f(y)).$$

Thus, (ii) implies (iii). That (iii) implies (i) and that (i) implies (ii) are immediate consequences of 1.14 and 1.15.

THEOREM 1.17. *If f is a function on A to A, then the following conditions are equivalent:*

(i) *f is additive, and $f(f(x)) = x\cdot f(1)$ for any $x \varepsilon A$.*

(ii) *$f(x)\cdot y = f(x\cdot f(y))$ for any $x, y \varepsilon A$.*

Proof. Suppose (i) holds. Then, by 1.1(iii),

$$f(0) \leq f(f(0)) = 0.$$

Hence f is normal by 1.1(i). We shall next show that

(1) $x\cdot y = 0$ implies that $f(x)\cdot f(y) = 0$ for any $x, y \varepsilon A$.

In fact, if $x\cdot y = 0$, then by (i) and 1.1(iii)

$$f(f(x)\cdot f(y)) \leq f(f(x))\cdot f(f(y)) = x\cdot y\cdot f(1) = 0,$$

and consequently

$$f(x)\cdot f(y) = f(x)\cdot f(y)\cdot f(1) = f(f(f(x)\cdot f(y))) = f(0) = 0.$$

Thus (1) holds. For any elements $x, y \varepsilon A$ we have, by (i) and 1.1(iii),

$$f(x)\cdot y = f(x)\cdot y\cdot f(1) = f(x)\cdot f(f(y))$$
$$= f(x)\cdot f(x\cdot f(y)) + f(x)\cdot f(x^{-}\cdot f(y)).$$

Hence, by (1) and 1.1(iii),

$$f(x)\cdot y = f(x\cdot f(y)) \text{ for any } x, y \varepsilon A,$$

and (ii) is satisfied. Conversely, if (ii) holds, then

$$f(0) = f(0\cdot f(0)) = f(0)\cdot 0 = 0,$$

so that f is normal by 1.1(i). Hence the first part of (i) follows from 1.16. The second part of (i) is but a special case of (ii).

THEOREM 1.18. *If f is a function on A to A, then the following conditions are equivalent:*

(i) *f is additive and $f(f(x)) = x$ for any $x \varepsilon A$.*

(ii) *$f(1) = 1$ and $f(x)\cdot y = f(x\cdot f(y))$ for any $x \varepsilon A$.*

Proof. Suppose (i) holds. Then $f(f(1)) = 1$ and hence, by 1.1(iii), $f(1) = 1$. Therefore, by 1.17, (ii) is satisfied. The converse is an immediate consequence of 1.17.

The representation theorem and the extension theorem for Boolean algebras were mentioned in the introduction, but for our later purposes the form in which they were given there is not strong enough. As regards the representation theorem, the stronger form which we have in mind is usually stated as follows:

Every Boolean algebra is isomorphic to a set-field consisting of all open and closed sets in a totally-disconnected compact space.[5]

Before stating the corresponding form of the extension theorem, we shall introduce the notions of a regular subalgebra and a perfect extension. The results contained in the last part of this section are substantially known, and the proofs will therefore be omitted.

DEFINITION 1.19. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1\rangle \text{ and } \mathfrak{B} = \langle B, +, 0, \cdot, 1\rangle$$

be two Boolean algebras. We say that $\mathfrak{B}$ *is a regular subalgebra of* $\mathfrak{A}$ *and that* $\mathfrak{A}$ *is a perfect extension of* $\mathfrak{B}$ *if the following conditions are satisfied:*

(i) $\mathfrak{A}$ *is complete and atomistic, and* $\mathfrak{B}$ *is a subalgebra of* $\mathfrak{A}$.

(ii) *If* I *is an arbitrary set, and if the elements* $x_i \varepsilon B$ *with* $i \varepsilon I$ *are such that*

$$\sum_{i \varepsilon I} x_i = 1,$$

then there exists a finite subset J *of* I *such that*

$$\sum_{i \varepsilon J} x_i = 1.$$

(iii) *If* u *and* v *are distinct atoms of* $\mathfrak{A}$, *then there exists an element* $b \varepsilon B$ *such that*

$$u \leq b \text{ and } v \cdot b = 0.$$

DEFINITION 1.20. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1\rangle$$

be a complete atomistic Boolean algebra, and let

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1\rangle$$

---

[5] Cf. Stone [1], Theorem 1. See the same work, Theorem 4, in connection with our Theorem 1.23.

be a regular subalgebra of $\mathfrak{A}$. *An element* $x \in A$ *is said to be*

(i) *open if* $x = \sum_{x \geq y \in B} y$;

(ii) *closed if* $x = \prod_{x \leq y \in B} y$.

THEOREM 1.21. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1 \rangle$$

be a complete and atomistic Boolean algebra, and let

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1 \rangle$$

be a regular subalgebra of $\mathfrak{A}$. *We then have:*

(i) *For any* $x \in A$, x *is open if, and only if,* x^{-} *is closed.*

(ii) *For any* $x \in A$, x *is open and closed if, and only if,* $x \in B$.

(iii) *If* $x \in A$ *is closed,* I *is an arbitrary set, the elements* $y_i \in A$ *with* $i \in I$ *are open, and*

$$x \leq \sum_{i \in I} y_i,$$

then there exists a finite subset J *of* I *such that*

$$x \leq \sum_{i \in J} y_i.$$

(iv) *If* $x \in A$ *is open,* I *is an arbitrary set, the elements* $y_i \in A$ *with* $i \in I$ *are closed, and*

$$x \geq \prod_{i \in I} y_i,$$

then there exists a finite subset J *of* I *such that*

$$x \geq \prod_{i \in J} y_i.$$

(v) *If* u *is an atom of* $\mathfrak{A}$, *then* u *is closed.*

(vi) *If* u *is an atom of* $\mathfrak{B}$, *then* u *is an atom of* $\mathfrak{A}$.

The extension theorem for Boolean algebras can now be stated as follows:

THEOREM 1.22. *For any Boolean algebra* $\mathfrak{B}$ *there exists a complete and atomistic Boolean algebra* $\mathfrak{A}$ *which is a perfect extension of* $\mathfrak{B}$.

That the perfect extension $\mathfrak{A}$ is essentially determined by the Boolean algebra $\mathfrak{B}$ is shown in the next two theorems.

THEOREM 1.23. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1 \rangle \text{ and } \mathfrak{A}' = \langle A', +, 0, \cdot, 1 \rangle$$

be two complete and atomistic Boolean algebras, and let

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1\rangle \text{ and } \mathfrak{B}' = \langle B', +, 0, \cdot, 1\rangle$$

be regular subalgebras of $\mathfrak{A}$ *and* $\mathfrak{A}'$ *respectively. If a function* ϕ *maps* $\mathfrak{B}$ *isomorphically onto* $\mathfrak{B}'$, *then there exists a unique function* ψ *which maps* $\mathfrak{A}$ *isomorphically onto* $\mathfrak{A}'$ *in such a way that* $\psi/B = \phi$.

THEOREM 1. 24. *If*

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1\rangle$$

is a Boolean algebra, and if $\mathfrak{A}$ *and* $\mathfrak{A}'$ *are two perfect extensions of* $\mathfrak{B}$, *then there exists a unique function* ψ *which maps* $\mathfrak{A}$ *isomorphically onto* $\mathfrak{A}'$ *in such a way that* $\psi(x) = x$ *for every* $x \varepsilon B$.

In the following we shall use various definitions and theorems from this section without always explicitly referring to them.

Section 2.

The Extension Theorem.

Throughout the first part of this section we shall consider a fixed Boolean algebra

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1\rangle$$

which is complete and atomistic, and a regular subalgebra

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1\rangle$$

of $\mathfrak{A}$. The set consisting of 0 and all the atoms of $\mathfrak{A}$ will be denoted by At, the set of all closed elements by C. Although we associate with every function f on B^m to B a function f^+ on A^m to A (see Definition 2. 1), we shall be mostly concerned with those functions f which are additive or obtainable from additive functions by means of composition. In this case f^+ is an extension of f. It is shown that if f is additive, then f^+ is completely additive. We shall also prove that if an equation involving additive functions on B^m to B is identically satisfied, then the corresponding equation involving their extensions will also hold identically. In the last part of the section these results are applied to Boolean algebras with operators.

DEFINITION 2. 1. *For any function f on B^m to B, f^+ is the function on A^m to A defined by the formula*

$$f^+(x) = \sum_{x \geq y \,\varepsilon\, C^m} \; \prod_{y \leq z \,\varepsilon\, B^m} f(z) \text{ for any } x \,\varepsilon\, A^m.$$

THEOREM 2. 2. *If f is a function on B^m to B, then*

$$f^+(y) = \prod_{y \leq z \,\varepsilon\, B^m} f(z) \text{ for any } y \,\varepsilon\, C^m.$$

Proof. By 2. 1.

THEOREM 2. 3. *If f is an additive (or, more generally, a monotonic) function on B^m to B, then*

$$f^+/B^m = f.$$

Proof. By 2. 2 with the aid of 1. 21(ii).

THEOREM 2. 4. *If f is an additive function on B^m to B, then f^+ is a completely additive function on A^m to A.*

Proof. Let g be the function on A^m to A such that

(1) $$g(x) = \sum_{x \geq u \,\varepsilon\, At^m} f^+(u) \text{ for any } x \,\varepsilon\, A^m.$$

By 1. 7, g is completely additive. In order to prove that $f^+ = g$, we shall first show that

(2) $$f^+(y) \leq g(y) \text{ for any } y \,\varepsilon\, C^m.$$

Consider a fixed sequence $y \,\varepsilon\, C^m$ and an atom v of $\mathfrak{A}$ such that

(3) $$v \cdot g(y) = 0.$$

We shall show that there exists a sequence z such that

(4) $$y \leq z \,\varepsilon\, B^m \text{ and } v \cdot f(z) = 0.$$

Let

(5) $$K = \underset{u}{\mathrm{E}}[y \geq u \,\varepsilon\, At^m].$$

For any sequence $u \,\varepsilon\, K$ we then have, by (1) and (3), $v \cdot f^+(u) = 0$. Hence, by 2. 2, we can correlate with each $u \,\varepsilon\, K$ a sequence u' such that

(6) $$u \leq u' \,\varepsilon\, B^m \text{ and } v \cdot f(u') = 0 \text{ whenever } u \,\varepsilon\, K.$$

We shall correlate with each sequence $u \,\varepsilon\, K$ elements $\phi_{j\,p}(u) \,\varepsilon\, B$ with $j < m$ and $p \leq m$ such that the following conditions are satisfied:

(7) $$y_j \leq \phi_{j\,p}(u) \text{ whenever } j < p \leq m \text{ and } u \,\varepsilon\, K.$$

(8) $u_j \leq \phi_{j\,p}(u)$ whenever $p \leq j < m$ and $u \in K$.

(9) $v \cdot f(\phi_{0\,p}(w), \phi_{1\,p}(u), \cdots, \phi_{m-1\,p}(u)) \doteq 0$ whenever $p \leq m$ and $u \in K$.

The elements $\phi_{jp}(u)$ will be defined recursively. We let

$$\phi_{j\,0}(u) = u'_j \text{ for any } j < m \text{ and } u \in K.$$

Assume that $p < m$, that $\phi_{j\,p}(u)$ has been defined for any $j < m$ and $u \in K$, and that (7), (8), and (9) are satisfied for this fixed value of p. Consider a fixed sequence $u \in K$ and the class L of all sequences $w \in K$ such that $w_j = u_j$ whenever $j < m$ and $j \neq p$. For each atom $a \leq y_p$ we have a sequence $w \in L$ such that $w_p = a$ and hence $a \leq \phi_{p\,p}(w)$. Consequently,

$$y_p \leq \sum_{w \in L} \phi_{p\,p}(w).$$

Since y_p is closed and $\phi_{p\,p}(w) \in B$, there exists, by 1.21(ii),(iii), a finite subset M of L such that

$$y_p \leq \sum_{w \in M} \phi_{p\,p}(w).$$

Let

$$\phi_{p\,p+1}(u) = \sum_{w \in M} \phi_{p\,p}(w),$$

$$\phi_{j\,p+1}(u) = \prod_{w \in M} \phi_{j\,p}(w) \text{ whenever } j < m \text{ and } j \neq p.$$

It is easy to see that (7), (8), and (9) are satisfied with p replaced by $p + 1$. We now take any sequence $u \in K$, define the m-termed sequence $z \in B^m$ by the formula

$$z_j = \phi_{j\,m}(u) \text{ for } j < m,$$

and use (7) and (9) to check that (4) is satisfied. Hence, by 2.2, $v \cdot f^+(y) = 0$. Since this is true whenever $y \in C^m$, v is an atom of $\mathfrak{A}$, and (3) is satisfied, we conclude that (2) holds.

By (1), (2), 2.1, and 2.2 we have, for any $x \in A^m$,

$$f^+(x) = \sum_{x \geq y \in C^m} f^+(y) \leq \sum_{x \geq y \in C^m} g(y) \leq g(x).$$

Thus $f^+ \leqq g$. The opposite inequality follows immediately from (1) and 2.1. Hence $f^+ = g$, and the proof is complete.

THEOREM 2.5. *If f is an additive function on B^m to B and g is a completely additive function on A^m to A such that $g/B^m \leq f$, then $g \leq f^+$.*

Proof. For any $u \in At^m$ we have, by 2.2 and the hypothesis,

$$f^+(u) = \prod_{u \leq y \in B^m} f(y) \geq \prod_{u \leq y \in B^m} g(y) \geq g(u).$$

Hence the conclusion follows by 1.8(i), 2.4, and the hypothesis.

It may be interesting to notice that, in general, if f is an additive function on B^m to B, then there exist, besides f^+, other completely additive functions g on A^m to A which satisfy the condition $g/B^m = f$. To give a simple example, let f be the function on B to B such that

$$f(0) = 0, \text{ and } f(x) = 1 \text{ whenever } 0 \neq x \in B.$$

It follows immediately from 2.1 that

$$f^+(0) = 0, \text{ and } f^+(x) = 1 \text{ whenever } 0 \neq x \in A.$$

Observe that, if B is infinite, then At is not a subset of B. For, assume that $At \subseteq B$. Since

$$\sum_{u \in At} u = 1,$$

there exists, by 1.19(ii), a finite subset K of At such that

$$\sum_{u \in K} u = 1.$$

But this clearly implies that K contains all the atoms of $\mathfrak{A}$. Hence At is finite, which in turn implies that A and B are finite. Assuming that B is infinite, we can therefore choose an atom u of $\mathfrak{A}$ such that $u \notin B$. Let the function g on A to A be defined by the formulas

$$g(0) = 0 = g(u), \text{ and } g(x) = 1 \text{ whenever } x \in A \text{ and } x \neq 0, u.$$

Obviously, g is completely additive, and we have $g/B = f$ while $g \neq f^+$.

If f is an additive function on B^m to B, and if g is a completely additive function on A^m to A such that $g/B^m = f$, then $g \leq f^+$ by 2.5. We may express this by saying that f^+ is the largest possible extension of f which is completely additive. However, if we drop the condition that the extension is completely additive and consider instead all additive extensions of f, then f^+ no longer has this property. In fact, the largest possible extension of f which is additive is the function g such that

$$g(x) = \prod_{x \leq y \in B^m} f(y) \text{ for any } x \in A^m,$$

and it is easy to show by examples that we do not in general have $g = f^+$. As a matter of fact, if B is infinite, and if f is an identity function on B^m to B, then $f^+ \neq g$. For, as we will prove in the next theorem, f^+ is then the corresponding identity function on A^m to A. Hence $f^+ = g$ implies that every element of A is closed and, therefore, by 1.21(i), (ii), that $A = B$. But we have already shown that this is true only when B is finite.

THEOREM 2.6. *If f and g are the functions on B^2 to B defined by the formulas*

$$f(x) = x_0 + x_1 \text{ and } g(x) = x_0 \cdot x_1 \text{ for any } x \in B^2,$$

then

$$f^+(x) = x_0 + x_1 \text{ and } g^+(x) = x_0 \cdot x_1 \text{ for any } x \in A^2.$$

Furthermore, if $j < m$ and $b \in B$, and if h and k are the functions on B^m to B defined by the formulas

$$h(x) = b \text{ and } k(x) = x_j \text{ for any } x \in B^m,$$

then

$$h^+(x) = b \text{ and } k^+(x) = x_j \text{ for any } x \in A^m.$$

Proof. For any $u \in At^2$ we have, by 2.2,

$$f^+(u) = \prod_{u \leq y \in B^2} (y_0 + y_1). \tag{1}$$

Therefore

$$u_0 + u_1 \leq f^+(u). \tag{2}$$

Consider any atom a of $\mathfrak{A}$ with $a \cdot u_0 = 0 = a \cdot u_1$. By 1.19(iii) there exists a sequence $y \in B^2$ such that $u \leq y$ and $a \cdot y_0 = 0 = a \cdot y_1$. It follows by (1) that $a \cdot f^+(u) = 0$ for every such atom a. Consequently $f^+(u) \leq u_0 + u_1$. Together with (2) this gives

$$f^+(u) = u_0 + u_1.$$

Since this formula holds whenever $u \in At^2$, we conclude by 1.8(ii) and 2.4 that

$$f^+(x) = x_0 + x_1 \text{ for any } x \in A^2.$$

To show that

$$g^+(x) = x_0 \cdot x_1 \text{ for any } x \in A^2,$$

we proceed in an analogous manner.

By 1.20(ii), 1.21(v) and 2.2 we have $k^+(u) = u_j$ for any $u \in At^m$. Hence, by 1.8(ii) and 2.4,

$$k^+(x) = x_j \text{ for any } x \in A^m.$$

By 2.1 we have

$$h^+(x) = b \text{ for any } x \in A^m.$$

The proof is complete.

THEOREM 2.7. *If f is a function on B^m to B and g is the function on B^{m+n} to B such that*

$$g(x^\frown y) = f(x) \textit{ for any } x \,\varepsilon\, B^m \textit{ and } y \,\varepsilon\, B^n,$$

then

$$g^+(x^\frown y) = f^+(x) \textit{ for any } x \,\varepsilon\, A^m \textit{ and } y \,\varepsilon\, A^n.$$

Proof. By 2.1.

THEOREM 2.8. *Let f be an additive function on B^m to B, and let $g_0, g_1, \cdots, g_{m-1}$ be additive functions on B^n to B. Then*

$$(f\langle g_0, g_1, \cdots, g_{m-1}\rangle)^+ = f^+\langle g^+{}_0, g^+{}_1, \cdots, g^+{}_{m-1}\rangle.$$

Proof. Let

$$(1) \qquad h = f\langle g_0, g_1, \cdots, g_{m-1}\rangle, \qquad k = f^+\langle g^+{}_0, g^+{}_1, \cdots, g^+{}_{m-1}\rangle.$$

Then h^+ and k are completely additive functions on A^{mn} to A by 1.9 and 2.4. In order to prove that $h^+ = k$, it is therefore sufficient to show that

$$(2) \qquad h^+(y) = k(y) \text{ for any } y \,\varepsilon\, C^{mn}.$$

For each sequence $y \,\varepsilon\, C^{mn}$ let y' be the m-termed sequence defined by the formula

$$(3) \qquad y'_j = g_j^+(y_{j\,n}, y_{j\,n+1}, \cdots, y_{j\,n+n-1}) \text{ for any } j < m.$$

If $y \,\varepsilon\, C^{mn}$, then $y' \,\varepsilon\, C^m$ by (3) and 2.2. Hence, by (1) and 2.2,

$$(4) \quad h^+(y) = \prod_{y \leq z \,\varepsilon\, B^{mn}} f(z') \text{ and } k(y) = \prod_{y' \leq t \,\varepsilon\, B^m} f(t) \text{ for any } y \,\varepsilon\, C^{mn}.$$

By (3), 2.3, and the additivity of $g_0, g_1, \cdots, g_{m-1}$ we have

$$y' \leqq z' \,\varepsilon\, B^m \text{ whenever } y \,\varepsilon\, C^{mn},\ z \,\varepsilon\, B^{mn}, \text{ and } y \leq z.$$

Consequently, by (4) and the additivity of f,

$$(5) \qquad k(y) \leq h^+(y) \text{ for any } y \,\varepsilon\, C^{mn}.$$

To prove the inequality in the opposite direction, it is clearly sufficient to show that if

$$(6) \qquad y \,\varepsilon\, C^{mn} \text{ and } y' \leq t \,\varepsilon\, B^m,$$

then there exists a sequence z such that

$$(7) \qquad y \leq z \,\varepsilon\, B^{mn} \text{ and } z' \leq t.$$

Assume that (6) holds and let

$$u^{(j)} = y_{j\,n}{}^\frown y_{j\,n+1}{}^\frown \cdots {}^\frown y_{j\,n+n-1} \text{ for any } j < m.$$

Then

$$\prod_{u^{(j)} \leq v \,\varepsilon\, B^n} g_j(v) = y'_j \leq t_j \text{ for any } j < m.$$

Since $t_j \varepsilon B$ by (6), and $g_j(v) \varepsilon B$ for $v \varepsilon B^n$, it follows by 1.21(ii)(iv) that there exists, for each $j < m$, a finite set K_j of sequences v with $u^{(j)} \leq v \varepsilon B^n$ such that

$$\prod_{v \varepsilon K_j} g_j(v) \leq t_j \text{ for any } j < m.$$

Let z be the mn-termed sequence defined by the formula

$$z_{jn+p} = \prod_{v \varepsilon K_j} v_p \text{ for any } j < m \text{ and } p < n.$$

It is easy to check that (7) is satisfied. Thus (6) always implies (7), and we conclude that the inclusion in (5) can be replaced by an equality. The proof is complete.

THEOREM 2.9. *Let f be a monotonic function on B^m to B, and let $g_0, g_1, \cdots, g_{m-1}$ be identity functions on B^n to B. Then*

$$(f[g_0, g_1, \cdots, g_{m-1}])^+ = f^+[g^+_0, g^+_1, \cdots, g^+_{m-1}].$$

Proof. Let

(1) $$h = f[g_0, g_1, \cdots, g_{m-1}].$$

For each $x \varepsilon A^n$, let x' be the m-termed sequence such that

(2) $$x'_j = g^+_j(x) \text{ for any } j < m.$$

We shall prove the following two statements which will be needed later:

(3) If $x \varepsilon A^n$ and $x' \geq z \varepsilon C^m$, then $x \geq y$ and $y' \geq z$ for some $y \varepsilon C^n$.

(4) If $x \varepsilon A^n$ and $x' \leq t \varepsilon B^m$, then $x \leq s$ and $s' \leq t$ for some $s \varepsilon B^n$.

For each $p \leq n$, let k_p be the function on B^n to B such that

(5) $$k_p(x) = x_p \text{ for any } x \varepsilon B^n,$$

and let

(6) $$K_p = \underset{j}{\mathrm{E}}[j < m \text{ and } g_j = k_p].$$

The sequences y and s are defined by the formulas

(7) $$y_p = \sum_{j \varepsilon K_p} z_j \text{ and } s_p = \prod_{j \varepsilon K_p} t_j \text{ for any } p < n.$$

From the finiteness of the sets K_p it follows that $y \varepsilon C^n$ and $t \varepsilon B^n$. If $j \varepsilon K_p$, then, by (2), (5), and (6),

$$x'_j = g^+_j(x) = k^+_p(x) = x_p.$$

Consequently, by our assumptions regarding z and t,

$$z_j \leq x_p \leq t_j \text{ whenever } p < n \text{ and } j \varepsilon K_p.$$

Therefore, by (7),

$$y_p \leq x_p \leq s_p \text{ for any } p < n. \tag{8}$$

If $j < m$, then $j \in K_p$ for some $p < n$. Hence, by (2), (5), (6), and (7),

$$y'_j = y_p \geq z_j \text{ and } s'_j = s_p \leq t_j.$$

We therefore have

$$y' \geq z \text{ and } s' \leq t.$$

Together with (8) this shows that the conclusions of (3) and (4) are satisfied.

Consider now a fixed sequence $x \in A^n$. By (1), (2), and 2.1 we have

$$h^+(x) = \sum_{x \geq y \in C^n} \prod_{y \leq s \in B^n} f(s'), \tag{9}$$

$$f^+(x') = \sum_{x' \geq z \in C^m} \prod_{z \leq t \in B^m} f(t). \tag{10}$$

For any sequence y with $x \geq y \in C^n$, we have, by (2), 2.1, and 2.2, $x' \geq y' \in C^m$. Furthermore, if $y' \leq t \in B^m$, then there exists, by (4), a sequence $s \in B^n$ such that $y \leq s, s' \leq t$, and hence $f(s') \leq f(t)$. We thus have

$$\prod_{y \leq s \in B^n} f(s') \leq \prod_{y' \leq t \in B^m} f(t).$$

Since this is true whenever $x \geq y \in C^n$, we conclude, by (9) and (10), that

$$h^+(x) \leq f^+(x'). \tag{11}$$

For any sequence z with $x' \geq z \in C^m$ there exists, by (3), a sequence $y \in C^n$ such that $x \geq y, y' \geq z$, and hence

$$\prod_{z \leq t \in B^m} f(t) \leq \prod_{y' \leq t \in B^m} f(t).$$

Furthermore, since $y \leq s \in B^n$ always implies that $y' \leq s' \in B^m$, we have

$$\prod_{y' \leq t \in B^m} f(t) \leq \prod_{y \leq s \in B^n} f(s').$$

For each sequence z with $x' \geq z \in C^m$ there exists, therefore, a sequence y such that $x \geq y \in C^n$ and

$$\prod_{z \leq t \in B^m} f(t) \leq \prod_{y \leq s \in B^n} f(s').$$

Consequently, by (9) and (10),

$$f^{+}(x') \leq k^{+}(x). \tag{12}$$

Since (11) and (12) hold for any $x \varepsilon A^n$, the conclusion of the theorem follows from (1) and (2).

THEOREM 2.10. *Let* Φ *be the smallest family of functions which satisfies the following conditions:*

(i) *If, for some* m, f *is an additive function on* B^m *to* B, *then* $f \varepsilon \Phi$.

(ii) *If, for some* m *and* n, $f \varepsilon \Phi$ *is a function on* B^m *to* B *while* $g_0, g_1, \cdots, g_{m-1} \varepsilon \Phi$ *are functions on* B^n *to* B, *then* $f[g_0, g_1, \cdots, g_{m-1}] \varepsilon \Phi$.

Given any function $f \varepsilon \Phi$ *on* B^m *to* B *and* m *functions* $g_0, g_1, \cdots, g_{m-1} \varepsilon \Phi$ *on* B^n *to* B, *we then have*

$$(f[g_0, g_1, \cdots, g_{m-1}])^{+} = f^{+}[g^{+}_0, g^{+}_1, \cdots, g^{+}_{m-1}].$$

Proof. By 1.10 there exist, for some p, an additive function f' on B^p to B, and identity functions $h_0, h_1, \cdots, h_{p-1}$ on B^m to B such that

$$f = f'[h_0, h_1, \cdots, h_{p-1}]. \tag{1}$$

Let

$$k_j = h_j[g_0, g_1, \cdots, g_{m-1}] \text{ for any } j < p. \tag{2}$$

Then each k_j is one of the functions $g_0, g_1, \cdots, g_{m-1}$, whence $k_j \varepsilon \Phi$. We can therefore represent k_j in the form

$$k_j = k'_j[l_0^{(j)}, l_1^{(j)}, \cdots, l_{q-1}^{(j)}] \tag{3}$$

where k'_j is an additive function on B^q to B, and $l_0^{(j)}, l_1^{(j)}, \cdots, l_{q-1}^{(j)}$ are identity functions on B^n to B. The assumption that we have the same number q for all values of j is justified by 1.3. By (1) and (2) we have

$$f[g_0, g_1, \cdots, g_{m-1}] = f'[k_0, k_1, \cdots, k_{p-1}]. \tag{4}$$

Let

$$g' = f'\langle k'_0, k'_1, \cdots, k'_{p-1}\rangle \text{ and } l = l^{(0)} \frown l^{(1)} \frown \cdots \frown l^{(p-1)}. \tag{5}$$

Then, by (3) and (4),

$$f[g_0, g_1, \cdots, g_{m-1}] = g'[l_0, l_1, \cdots, l_{mp-1}].$$

Consequently, by 2.9,

$$(f[g_0, g_1, \cdots, g_{m-1}])^+ = g'^+[l^+_0, l^+_1, \cdots, l^+_{mp-1}]. \tag{6}$$

By (1) and 2.9 we have

$$f^+ = f'^+[h^+_0, h^+_1, \cdots, h^+_{p-1}]$$

while, by 2.6,

$$k^+_j = h^+_j[g^+_0, g^+_1, \cdots, g^+_{m-1}] \text{ for any } j < p.$$

Consequently

$$f^+[g^+_0, g^+_1, \cdots, g^+_{m-1}] = f'^+[k^+_0, k^+_1, \cdots, k^+_{p-1}]. \tag{7}$$

From (3) and 2.9 it follows that

$$k^+_j = k'^+_j[l_0^{(j)+}, l_1^{(j)+}, \cdots, l_{q-1}^{(j)+}] \text{ for any } j < p \tag{8}$$

while, by (5) and 2.8,

$$g'^+ = f'^+\langle k'^+_0, k'^+_1, \cdots, k'^+_{p-1}\rangle.$$

Hence, by (5), (7), and (8),

$$f^+[g^+_0, g^+_1, \cdots, g^+_{m-1}] = g'^+[l^+_0, l^+_1, \cdots, l^+_{mp-1}]. \tag{9}$$

The conclusion follows from (6) and (9).

It follows from the preceding theorem that if an equation involving additive functions $f_0, f_1, \cdots, f_{m-1}$ on B^n to B holds for every sequence $x \varepsilon B^n$, then the corresponding equation with the functions $f^+_0, f^+_1, \cdots, f^+_{m-1}$ holds for every sequence $x \varepsilon A^n$. It would be natural to ask at this point what other properties are preserved by the extension. We have not made a detailed study of this question, and shall only make a few remarks here.

We first notice that a similar result can be obtained for certain implications between two equations, as well as between an inequality and an equation. In fact, we have

THEOREM 2.11. *Let Φ be as in Theorem 2.10, and let $f, g, h \varepsilon \Phi$ be functions on B^m to B.*

(i) *If*

$$f(x) = 0 \text{ implies that } g(x) = h(x) \text{ for every } x \varepsilon B^m,$$

then

$$f^+(x) = 0 \textit{ implies that } g^+(x) = h^+(x) \textit{ for every } x \in A^m.$$

(ii) *If*

$$f(x) \neq 0 \textit{ implies that } g(x) = h(x) \textit{ for every } x \in B^m,$$

then

$$f^+(x) \neq 0 \textit{ implies that } g^+(x) = h^+(x) \textit{ for every } x \in A^m.$$

Proof. Let k be the function on B to B such that

$$k(0) = 0, \text{ and } k(x) = 1 \text{ whenever } 0 \neq x \in B.$$

By the hypothesis of (i) we then have

$$(1) \qquad k(f(x)) + g(x) = k(f(x)) + h(x) \text{ for every } x \in B^m.$$

Consequently, the function k being additive, we get by 2.6 and 2.10

$$(2) \qquad k^+(f^+(x)) + g^+(x) = k^+(f^+(x)) + h^+(x) \text{ for every } x \in A^m.$$

By 2.1 and the definition of k we have

$$k^+(0) = 0, \text{ and } k^+(x) = 1 \text{ whenever } 0 \neq x \in A.$$

Hence, by (2), we obtain the conclusion of (i):

$$f^+(x) = 0 \text{ implies that } g^+(x) = h^+(x) \text{ for every } x \in A^m.$$

The proof of (ii) is analogous; we change $+$ to $\cdot$ in (1) and (2).

As a corollary of the preceding theorem we obtain

THEOREM 2.12. *Let f and g be functions on B to B. If f and g are conjugate, then f^+ and g^+ are conjugate.*

Proof. By 1.11, 1.14, and 2.11.

Theorem 2.11 can be generalized by considering implications in which the hypothesis is, not an equation of the form $f(x) = 0$ or an inequality of the form $f(x) \neq 0$, but a conjunction or disjunction of formulas of both these types. To obtain this generalization, we argue as in the proof of 2.11 and, by applying the function k defined there, we show that every implication of the kind considered can be equivalently replaced by a single equation or a conjunction of equations. For instance, the implication

$$\text{if } f_0(x) = 0 \text{ and } f_1(x) \neq 0, \text{ then } g(x) = h(x)$$

proves to be equivalent to the equation

$$k(f_0(x)) + k(f_1(x)) \cdot g(x) = k(f_0(x)) + k(f_1(x)) \cdot h(x).$$

On the other hand, we can give an example of a simple formula which is identically satisfied by an additive function f in a certain Boolean algebra $\mathfrak{B} = \langle B, +, 0, \cdot, 1 \rangle$, but which is not satisfied by any additive function g (whether an extension of f or not) in $\mathfrak{A} = \langle A, +, 0, \cdot, 1 \rangle$; $\mathfrak{A}$ is assumed to be, not necessarily a perfect extension of $\mathfrak{B}$, but an arbitrary complete Boolean algebra of which $\mathfrak{B}$ is a subalgebra. In fact, let U be the set of all positive integers, and let B be the family consisting of all finite subsets of U and of their complements (with respect to U). Suppose that $+$ and $\cdot$ are set-addition and set-multiplication, 0 is the empty set, and 1 is the set U. For every x in B, let $f(x)$ be the set consisting of the integer 2 and of all integers $n + 2$ where n is in x. As is easily seen, f is an additive function on B to B satisfying the condition:

$$f(x) \neq x \text{ for every } x \varepsilon B.$$

This condition implicitly involves, in addition to f, another additive function, namely the identity function on B to B, but it implies the following condition which involves only f:

$$f(f(x)) \neq f(x) \text{ for every } x \varepsilon B.$$

On the other hand there is no additive function g in $\mathfrak{A}$ which satisfies the corresponding condition

$$g(g(x)) \neq g(x) \text{ for every } x \varepsilon A.$$

For, g being any additive (or, more generally, monotonic) function on A to A, we obtain an element $c \varepsilon A$ for which $g(c) = c$ and hence $g(g(c)) = g(c)$ by putting

$$c = \sum_{x \leq g(x)} x.^{6}$$

One might ask whether Theorem 2.10 could be generalized by replacing Φ by some larger family of functions; in particular, it would be natural to consider the family of all monotonic functions. However, it can be shown that there exist monotonic functions f and g in one variable such that $(f[g])^{+} \neq f^{+}[g^{+}]$. In fact let

---

[6] Cf. Tarski [1].

$$\mathfrak{D} = \langle D, +, 0, \cdot, 1\rangle \text{ and } \mathfrak{E} = \langle E, +, 0, \cdot, 1\rangle$$

be Boolean algebras such that $\mathfrak{D}$ is infinite and complete, and $\mathfrak{E}$ is a perfect extension of $\mathfrak{D}$. Let U be the set of all the atoms of $\mathfrak{E}$, let B be the set of all functions on U to D, and let θ and I be the functions on U to D such that

$$\Theta(u) = 0 \text{ and } \mathrm{I}(u) = 1 \text{ for any } u \in U.$$

Suppose

$$\mathfrak{A} = \langle A, +, \Theta, \cdot, \mathrm{I}\rangle$$

is a perfect extension of the Boolean algebra

$$\mathfrak{B} = \langle B, +, \Theta, \cdot, \mathrm{I}\rangle.$$

We now take a fixed element $w \in U$ and define the functions f and g on B to B by the formulas

$$f(\alpha) = \Theta \text{ whenever } \alpha \in B \text{ and } \alpha(w) \neq 1,$$

$$f(\alpha) = \mathrm{I} \text{ whenever } \alpha \in B \text{ and } \alpha(w) = 1,$$

$$g(\alpha)(w) = \sum_{u \in U}{}' \alpha(u), \text{ and } g(\alpha)(v) = 0 \text{ whenever } \alpha \in B \text{ and } w \neq v \in U.$$

Here $\sum'$ refers to the sum in the Boolean algebra $\mathfrak{D}$. Clearly g is additive and f is monotonic. It will be shown that $(f[g])^+ \neq f^+[g^+]$.

For each $u \in U$ let

$$K_u = \mathop{\mathrm{E}}_{\beta}[\beta \in B \text{ and } w \leq \beta(u)], \qquad L_u = \mathop{\mathrm{E}}_{\beta}[\beta \in B \text{ and } u \leq \beta(w)],$$

$$\kappa_u = \prod_{\beta \in K_u} \beta, \qquad \lambda_u = \prod_{\beta \in L_u} \beta.$$

Furthermore, for each $u \in U$, let γ_u be the function on U to D such that

$$\gamma_u(u) = 1, \text{ and } \gamma_u(v) = 0 \text{ whenever } u \neq v \in U.$$

Finally, let

$$\xi = \sum_{u \in U} \kappa_u \text{ and } \eta = \sum_{u \in U} \gamma_u.$$

It is easy to see that

$$g^+(\kappa_u) = \lambda_u \text{ for any } u \in U$$

and hence

$$g^+(\xi) = \gamma_w.$$

Since $\gamma_w \varepsilon B$ and $\gamma_w(w) = 1$, it follows that

$$f^+(g^+(\xi)) = \mathrm{I}.$$

Letting

$$h = f[g],$$

we are going to show that $h^+(\xi) = \odot$. In fact, for any closed element $\zeta \leq \xi$ we have $\zeta \leq \eta$ and, therefore,

$$\zeta \leq \sum_{u \varepsilon M} \gamma_u$$

for some finite subset M of U. Thus

$$\zeta \leq \xi \cdot \sum_{u \varepsilon M} \gamma_u = \sum_{u \varepsilon M} \kappa_u.$$

Since D was assumed to be infinite, there exists an element $a \varepsilon D$ such that $a \neq 1$ and $u \leq a$ for every $u \varepsilon M$. If δ is the function on U to D such that

$$\delta(u) = a \text{ for any } u \varepsilon U,$$

then

$$h^+(\zeta) \leq h(\delta) = \odot.$$

Thus $h^+(\zeta) = \odot$ whenever $\zeta \leq \xi$ and ζ is closed. Consequently $h^+(\xi) = \odot$.

The results of this section can be generalized in another direction, namely by considering functions on one Boolean algebra into another Boolean algebra. Given two complete and atomistic Boolean algebras

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1 \rangle \text{ and } \mathfrak{A}' = \langle A', +, 0, \cdot, 1 \rangle,$$

and regular subalgebras

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1 \rangle \text{ and } \mathfrak{B}' = \langle B' +, 0, \cdot, 1 \rangle$$

of $\mathfrak{A}$ and $\mathfrak{A}'$ respectively, we can associate with any function f on B^m to B' a function f^+ on A to A' defined by the formula in Definition 2.1. The changes which have to be made in the formulation of Theorems 2.2-2.7 and 2.9 are quite obvious, and the proofs require no new ideas. The generalization of Theorem 2.8 involves six Boolean algebras, three complete and atomistic Boolean algebras $\mathfrak{A}$, $\mathfrak{A}'$, and $\mathfrak{A}''$ and regular subalgebras $\mathfrak{B}$, $\mathfrak{B}'$, and $\mathfrak{B}''$ of $\mathfrak{A}$, $\mathfrak{A}'$ and $\mathfrak{A}''$, respectively. The generalization of Theorem 2.10 is perhaps less obvious. Instead of the class Φ of functions described there we consider, for any Boolean algebras $\mathfrak{B} = \langle B, +, 0, \cdot, 1 \rangle$ and

$\mathfrak{B}' = \langle B', +, 0, \cdot, 1\rangle$ and for any non-negative integer m, the class $\Phi_m(\mathfrak{B}, \mathfrak{B}')$ consisting of all functions f on B^m to B which are of the form $f = g[h_0, h_1, \cdots, h_{n-1}]$ where n is a non-negative integer, $h_0, h_1, \cdots, h_{n-1}$ are identity functions on B^m to B, and g is an additive function on B^m to B'. We then have:

Let $\mathfrak{A}$, $\mathfrak{A}'$, and $\mathfrak{A}''$ be perfect extensions of the Boolean algebras $\mathfrak{B}$, $\mathfrak{B}'$ and $\mathfrak{B}''$, respectively. If $g_0, g_1, \cdots, g_{n-1} \in \Phi_m(\mathfrak{B}, \mathfrak{B}')$ and $f \in \Phi_n(\mathfrak{B}', \mathfrak{B}'')$, then $f[g_0, g_1, \cdots, g_{n-1}] \in \Phi_m(\mathfrak{B}, \mathfrak{B}'')$ and

$$(f[g_0, g_1, \cdots, g_{m-1}])^+ = f^+[g^+_0, g^+_1, \cdots, g^+_{m-1}].$$

Observe that Theorems 1. 23 and 1. 24, as well as Theorems 2. 16 and 2. 17 that follow, are easy consequences of this result.

The generalization of Theorem 2. 11 now offers no difficulty; the functions f, g, and h involved are assumed to belong to $\Phi_m(\mathfrak{B}, \mathfrak{B}')$. In order to generalize Theorem 2. 12 we must first extend the notion of conjugate functions. A function g on B' to B will be called a conjugate of the function f on B to B' if, for every $x \in B$ and $y \in B'$, the formulas

$$f(x) \cdot y = 0 \text{ and } g(y) \cdot x = 0$$

are equivalent. Theorem 2. 12 can then be extended to conjugate functions in this new sense.

It is interesting to notice that in the proofs of some of the theorems of this section—as well as in the proofs of their generalizations just indicated—the use of condition (iii) of Definition 1. 19 can be avoided. This leads to the following results:

Consider two complete and atomistic Boolean algebras

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1\rangle \text{ and } \mathfrak{A}' = \langle A', +, 0, \cdot, 1\rangle$$

and let

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1\rangle \text{ and } \mathfrak{B}' = \langle B', +, 0, \cdot, 1\rangle$$

be subalgebras of $\mathfrak{A}$ and $\mathfrak{A}'$ respectively. We then have:

(i) If f is an additive function on B^m to B, then there exists an additive function g on A^m to A such that $g/B^m = f$.

(ii) If $\mathfrak{A}$ and $\mathfrak{B}$ satisfy condition (ii) of Definition 1. 19, and if f is an additive function on B^m to B, then there exists a completely additive function g on A^m to B such that $g/B^m = f$.

To prove (ii) we define g as in the proof of Theorem 2.4. Then g is clearly a completely additive function with $g/B^m \leq f$. In order to show that $f \leq g/B^m$, we repeat with certain obvious changes the argument used in establishing formula (2) in the proof of Theorem 2.4.

Part (i) follows easily from (ii). In fact, there exists a complete and atomistic Boolean algebra $\mathfrak{A}_1 = \langle A_1, +, 0, \cdot, 1 \rangle$ which is a perfect extension of $\mathfrak{A}$. Then condition (ii) of Definition 1.19 holds if we understand $\sum$ to denote summation over $\mathfrak{A}_1$, and it follows from (ii) that there exists a completely additive function h on $A_1{}^m$ to A' such that $h/B^m = f$. Consequently the function $g = h/A^m$ is additive and we have $g/B^m = f$.

One might ask what properties of the function f beside its additivity are preserved by the extension considered in (i) and in (ii). We have not studied this problem in detail, and shall only remark that if in (ii) we assume that $\mathfrak{A}$ is a perfect extension of $\mathfrak{B}$ and that f maps $\mathfrak{B}$ homomorphically into $\mathfrak{B}'$, then it can be shown that g maps $\mathfrak{A}$ homomorphically into $\mathfrak{A}'$. However, we omit the proof of this statement, since a much more general theorem of the same type is known from the literature.[7]

To conclude this section, we define formally the notion of a Boolean algebra with operators and we extend the notion of a perfect extension to this kind of algebras (Definitions 2.13 and 2.14). Using these new notions, we shall be able to summarize the main results so far achieved in a concise and convenient manner (Theorems 2.15 and 2.18).

DEFINITION 2.13. *By a Boolean algebra with operators we shall mean an algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, f_0, f_1, \cdots, f_\xi, \cdots \rangle$$

such that $\langle A, +, 0, \cdot, 1 \rangle$ *is a Boolean algebra and the functions* f_ξ *are additive.*

Suppose $\mathfrak{A}$ *is a Boolean algebra with operators. By an atom of* $\mathfrak{A}$ *we mean an atom of the Boolean algebra* $\langle A, +, 0, \cdot, 1 \rangle$. *We say that* $\mathfrak{A}$ *is atomistic if the Boolean algebra* $\langle A, +, 0, \cdot, 1 \rangle$ *is atomistic. We call* $\mathfrak{A}$ *complete if the Boolean algebra* $\langle A, +, 0, \cdot, 1 \rangle$ *is complete and if each of the operations* f_ξ *is completely additive. Finally, we say that* $\mathfrak{A}$ *is normal if each of the operations* f_ξ *is normal.*

DEFINITION 2.14. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, f_0, f_1, \cdots, f_\xi, \cdots \rangle$$

---

[7] Cf. Sikorski [1], p. 332. Related problems are discussed in Sikorski [2].

and

$$\mathfrak{B} = \langle B, 0, +, \cdot, 1, g_0, g_1, \cdots, g_\xi, \cdots \rangle$$

be two Boolean algebras with operators. We say that $\mathfrak{A}$ *is a perfect extension of* $\mathfrak{B}$ *and that* $\mathfrak{B}$ *is a regular subalgebra of* $\mathfrak{A}$ *if the following conditions are satisfied:*

(i) $\mathfrak{A}$ *is complete and atomistic and* $\mathfrak{B}$ *is a subalgebra of* $\mathfrak{A}$.

(ii) $\langle B, +, 0, \cdot, 1 \rangle$ *is a regular subalgebra of* $\langle A, +, 0, \cdot, 1 \rangle$.

(iii) *For each of the functions* f_ξ *we have*

$$f_\xi(x) = \prod_{x \leqq y \in B^{m_\xi}} g_\xi(y) \text{ whenever } x \in At^{m_\xi},$$

where m_ξ *is the rank of* f_ξ *and* At *has its usual meaning.*

If $\mathfrak{A}$ is a perfect extension of $\mathfrak{B}$, we clearly have $f_\xi = g^+{}_\xi$.

THEOREM 2.15 (EXTENSION THEOREM). *For any Boolean algebra with operators* $\mathfrak{B}$ *there exists a complete and atomistic Boolean algebra with operators* $\mathfrak{A}$ *which is a perfect extension of* $\mathfrak{B}$.

Proof. Let

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1, g_0, g_1, \cdots, g_\xi, \cdots \rangle$$

be a Boolean algebra with operators. By 1.22 the Boolean algebra $\langle B, +, 0, \cdot, 1 \rangle$ is a regular subalgebra of a complete and atomistic Boolean algebra $\langle A, +, 0, \cdot, 1 \rangle$. If we put $f_\xi = g^+{}_\xi$ for each ξ, then, by 2.3, $\mathfrak{B}$ is a subalgebra of the algebra

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, f_0, f_1, \cdots, f_\xi, \cdots \rangle.$$

Moreover, by 2.4 and 2.13, $\mathfrak{A}$ is a complete and atomistic Boolean algebra with operators. Finally, 2.14(iii) is satisfied by 2.2. Hence $\mathfrak{A}$ is a perfect extension of $\mathfrak{B}$.

Just as in the case of ordinary Boolean algebras, the perfect extension $\mathfrak{A}$ is essentially determined by the Boolean algebra with operators $\mathfrak{B}$. This is shown in the next two theorems.

THEOREM 2.16. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, f_0, f_1, \cdots, f_\xi, \cdots \rangle$$

and

$$\mathfrak{A}' = \langle A', +, 0, \cdot, 1, f'_0, f'_1, \cdots, f'_\xi, \cdots \rangle$$

be two complete and atomistic Boolean algebras with operators, and let

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1, g_0, g_1, \cdots, g_\xi, \cdots \rangle$$

and

$$\mathfrak{B}' = \langle A', +, 0, \cdot, 1, g'_0, g'_1, \cdots, g'_\xi, \cdots \rangle$$

be regular subalgebras of $\mathfrak{A}$ *and* $\mathfrak{A}'$*, respectively. If a function* ϕ *maps* $\mathfrak{B}$ *isomorphically onto* $\mathfrak{B}'$*, then there exists a unique function* ψ *which maps* $\mathfrak{A}$ *isomorphically onto* $\mathfrak{A}'$ *in such a way that* $\psi/B = \phi$.

Proof. By 1.23 there exists a unique function ψ such that

$$(1) \qquad \psi/B = \phi \text{ and } \langle A, +, 0, \cdot, 1\rangle \underset{\psi}{\cong} \langle A', +, 0, \cdot, 1\rangle.$$

Let At be the set consisting of 0 and all the atoms of $\mathfrak{A}$. By (1) we have

$$(2) \qquad \psi(x) = \sum_{x \geq u \,\varepsilon\, At} \; \prod_{u \leq y \,\varepsilon\, B} \phi(y) \text{ for any } x \,\varepsilon\, A.$$

Consider any operation f_ξ. If the rank of f_ξ is m, then, by 2.14,

$$f_\xi(x) = \sum_{x \geq u \,\varepsilon\, At^m} \; \prod_{u \leq y \,\varepsilon\, B^m} g_\xi(y) \text{ for any } x \,\varepsilon\, A^m.$$

Therefore, by (1) and (2),

$$\psi(f_\xi(x)) = \sum_{x \geq u \,\varepsilon\, At^m} \; \prod_{u \leq y \,\varepsilon\, B^m} g'_\xi(\phi(y_0), \phi(y_1), \cdots, \phi(y_{m-1}))$$

for any $x \,\varepsilon\, A^m$, and we easily conclude by (1) and 2.14 that

$$\psi(f_\xi(x)) = f'_\xi(\psi(x_0), \psi(x_1), \cdots, \psi(x_{m-1})) \text{ for any } x \,\varepsilon\, A^m.$$

Consequently,

$$\mathfrak{A} \underset{\psi}{\cong} \mathfrak{A}'.$$

THEOREM 2.17. *If*

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1, g_0, g_1, \cdots, g_\xi, \cdots \rangle$$

is a Boolean algebra with operators, and if $\mathfrak{A}$ *and* $\mathfrak{A}'$ *are two perfect extensions of* $\mathfrak{B}$*, then there exists a unique function* ψ *mapping* $\mathfrak{A}$ *isomorphically onto* $\mathfrak{A}'$ *in such a way that* $\psi(x) = x$ *for every* $x \,\varepsilon\, B$.

Proof. By 2.16.

THEOREM 2.18. *Let*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, f_0, f_1, \cdots, f_\xi, \cdots \rangle$$

and

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1, g_0, g_1, \cdots, g_\xi, \cdots \rangle$$

be Boolean algebras with operators such that $\mathfrak{A}$ *is a perfect extension of* $\mathfrak{B}$. *If* L *is any equationally definable class of algebras such that* $\mathfrak{B}$ *belongs to* L, *then* $\mathfrak{A}$ *also belongs to* L.

Proof. By 2.2, 2.4, and 2.14 we have $f_\xi = g^+{}_\xi$ for each ξ. By 2.6, the same relation holds between the corresponding Boolean operations in $\mathfrak{A}$ and $\mathfrak{B}$; since $\mathfrak{B}$ is a subalgebra of $\mathfrak{A}$, this also applies to 0 and 1 (if they are treated as operations with rank 0). If now h is an algebraic function over the similarity class to which $\mathfrak{A}$ and $\mathfrak{B}$ belong, then $h_{\mathfrak{A}}$ can be obtained by means of a definite process of iterated composition from the fundamental operations of $\mathfrak{A}$, and $h_{\mathfrak{B}}$ can be obtained by means of the same process from the corresponding operations of $\mathfrak{B}$. Hence, by 2.10.

$$h_{\mathfrak{A}} = h^+{}_{\mathfrak{B}}.$$

Similarly, if k is another algebraic function over the same similiarity class, we have

$$k_{\mathfrak{A}} = k^+{}_{\mathfrak{B}}.$$

Consequently, $h_{\mathfrak{B}} = k_{\mathfrak{B}}$ implies $h_{\mathfrak{A}} = k_{\mathfrak{A}}$; i. e., every equation between algebraic functions which holds (identically) in $\mathfrak{B}$ holds in $\mathfrak{A}$ as well. (For notions involved in this theorem and in its proof see Introduction.)

Section 3.

The Representation Theorem; Application to Closure Algebras and Cylindric Algebras.

A set-field B whose universal set is U will be called *regular* if the Boolean algebra $\langle \mathsf{B}, \cup, \Lambda, \cap, U \rangle$ is a regular subalgebra of the Boolean algebra $\langle \mathsf{A}, \cup, \Lambda, \cap, U \rangle$, where A is the family of all subsets of U. The representation theorem for Boolean algebras can then be stated as follows:

THEOREM 3.1. *Given any Boolean algebra*

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1 \rangle,$$

there exists a regular set-field B' *such that* $\mathfrak{B}$ *is isomorphic to the Boolean algebra*

$$\mathfrak{B}' = \langle \mathsf{B}', \cup, \Lambda, \cap, U \rangle,$$

where U *is the universal set of* B'.

As a consequence of this theorem, the study of Boolean algebras reduces to the study of regular set-fields. We shall therefore proceed to investigate additive functions over regular set-fields. This will lead us to a representation theorem for Boolean algebras with operators (Theorem 3. 10). In the second part of this section we apply the extension theorem and the representation theorem to closure algebras and cylindric algebras.

Throughout the first part of this section we fix a regular set-field B, let U be the universal set of B, and let A be the family of all subsets of U. In accordance with our previous convention, At will denote the set consisting of the zero element and all the atoms of the Boolean algebra $\langle \mathsf{A}, \cap, \Lambda, \cup, U \rangle$, i. e., At consists of the empty set Λ and all sets of the form $\{x\}$ with $x \in U$.

We begin by establishing a one-to-one correspondence between all normal and completely additive functions on A^m to A and all $(m+1)$-ary relations $R \subseteq U^{m+1}$.

DEFINITION 3. 2. *If* $R \subseteq U^{m+1}$, *then the function* R^* *on* A^m *to* A *is defined by the formula*

$$R^*(X) = \mathop{\mathrm{E}}_{y} [x^\frown y \in R \text{ for some } x \in X_0 \times X_1 \times \cdots \times X_{m-1}] \text{ for any } X \in \mathsf{A}^m.$$

In this definition the last terms of the sequences belonging to R play a special role. This may be somewhat unnatural in case R is not a function and we would of course correlate with R m other functions on A^m to A besides R^*. However, these functions will not be needed in the present work.

THEOREM 3. 3. *If* $R \subseteq U^{m+1}$, *then* R^* *is a normal and completely additive function on* A^m *to* A.

Conversely, if F *is a normal and completely additive function on* A^m *to* A, *then there exists a unique relation* $R \subseteq U^{m+1}$ *such that* $F = R^*$. *This relation is defined by the formula*

$$R = \mathop{\mathrm{E}}_{x^\frown y} [x \in U^m \text{ and } y \in F(\{x_0\}, \{x_1\}, \cdots, \{x_{m-1}\})].^{8}$$

---

[8] Theorem 3.3 (restricted to functions of rank 1) is stated without proof in Tarski [3]. The same applies to Theorem 3.6 and to that part of Theorem 3.5 which concerns the equivalence of (i_4) and (ii_4).

Proof. The first part follows from 1.6(i′)(ii′) and 3.2. To prove the second part, suppose F is a normal and completely additive function on A^m to A, and let R be defined as above. Since R^* is a normal and completely additive function on A^m to A by the first part of the theorem, we see by 1.8(ii) that $F = R^*$ is equivalent to the formula

$$(1) \qquad F(X) = R^*(X) \text{ whenever } X \varepsilon \mathsf{At}^m.$$

If $X_j = \Lambda$ for some $j < m$, then this follows from the normality of F and R^*. If $X_j \neq \Lambda$ for every $j < m$, then there exists a sequence $x \varepsilon U^m$ such that

$$X_j = \{x_j\} \text{ for every } j < m.$$

For any element $y \varepsilon U$ we then have

$$y \varepsilon F(X) \text{ if and only if } x^\frown y \varepsilon R.$$

By 3.2 we have

$$x^\frown y \varepsilon R \text{ if and only if } y \varepsilon R^*(X),$$

and we conclude that (1) holds in this case also. The uniqueness of R is an immediate consequence of 3.2.

THEOREM 3.4. *If F is a normal and additive function on B^m to B, then there exists a relation $R \subseteq U^{m+1}$ such that $F = R^*/\mathsf{B}^m$.*

Proof. By 2.3, 2.4, and the hypothesis, F^+ is a normal and completely additive function with $F = F^+/\mathsf{B}^m$. Hence, by 3.3, there exists a relation $R \subseteq U^{m+1}$ such that $F^+ = R^*$ and, consequently, $F = R^*/\mathsf{B}^m$.

It follows from 3.3 and 3.4 that given a normal and completely additive function F on A^m to A, or a normal and additive function F on B^m to B, we can express any property that F may possess as a property of the corresponding relation R. In the next three theorems we provide some examples of how this can be done. In order to state these theorems, we shall need certain familiar notions from the calculus of binary relations. The *domain* of a binary relation R is the set consisting of all elements x such that $\langle x, y\rangle \varepsilon R$ for some element y; the *field* of R is the union of the domain of R and the domain of R^{-1}. A binary relation R is said to be *reflexive* if $\langle x, x\rangle \varepsilon R$ whenever x is in the field of R, *symmetric* if $R^{-1} \subseteq R$, and *transitive* if $R|R \subseteq R$. If R is both symmetric and transitive (and hence also reflexive), then we call it an *equivalence relation.*

THEOREM 3.5. *Let F be a function on A to A. Assume that F is normal and completely additive and satisfies one or more of the following conditions:*

(i_1) $X \subseteq F(X)$ *for any* $X \varepsilon \mathsf{A}$,

(i_2) $F(F(X)) \subseteq F(X)$ *for any* $x \varepsilon \mathsf{A}$,

(i_3) F *is selfconjugate*,

(i_4) $F(X \cap Y) = F(X) \cap F(Y)$ *for any* $X, Y \varepsilon \mathsf{A}$,

(i_5) $F(F(X)) = X \cap F(U)$ *for any* $X \varepsilon \mathsf{A}$,

(i_6) $F(F(X)) = X$ *for any* $X \varepsilon \mathsf{A}$.

Then there exists a relation $R \subseteq U^2$ *with* $F = R^*$ *which satisfies the corresponding condition, or conditions, in the following list:*

(ii_1) R is *reflexive and the domain of* R *is* U,

(ii_2) R *is transitive*,

(ii_3) R *is symmetric*,

(ii_4) R^{-1} *is a function*,

(ii_5) R *is symmetric and is a function*,

(ii_6) R *is symmetric and is a function whose domain is* U.

Conversely, if there is a relation $R \subseteq U^2$ *with* $F = R^*$ *which satisfies one or more of the conditions* (ii_1)-(ii_6), *then* F *is normal and completely additive and satisfies the corresponding conditions* (i_1)-(i_6).

Proof. By 3.3, F is normal and completely additive if and only if $F = R^*$ for some $R \subseteq U^2$. Assuming that $F = R^*$ we shall prove that the conditions (i_5) and (ii_5) are equivalent (assuming that the equivalence of (i_3) and (ii_3) has already been established); the corresponding proofs for the remaining conditions will be omitted.

Assume that (i_5) holds. For any element $x \varepsilon U$ we then have

(1) $F(F(\{x\})) = \{x\}$ if $x \varepsilon F(U)$, and $F(F(\{x\})) = \Lambda$ if $x \not\varepsilon F(U)$.

Consider two elements $x, y \varepsilon U$ with $\langle x, y \rangle \varepsilon R$. Then, by 3.2, $y \varepsilon F(\{x\})$ whence, by the additivity of F, $y \varepsilon F(U)$. By (1) we therefore have $F(F(\{y\})) = \{y\}$ and, F being normal, this implies that $F(y) \neq \Lambda$. Consequently, by the additivity of F, $F(F(\{x\})) \neq \Lambda$. By (1) this implies that $F(F\{x\})) = \{x\}$ and hence $F(\{y\}) = \{x\}$. We therefore see by 3.2 that $\langle y, x \rangle \varepsilon R$. Thus R is symmetric. To show that R is a function, consider two elements $x, y \varepsilon U$ such that $\langle y, x \rangle \varepsilon R$. Since R is symmetric, we have $\langle x, y \rangle \varepsilon R$; as we have already shown, this leads to the conclusion $F(\{y\}) = \{x\}$. Together with 3.2 this shows that for each element $y \varepsilon U$ there is at most one element $x \varepsilon U$ such that $\langle y, x \rangle \varepsilon R$. Hence R is a function and (ii_5) holds.

If (ii_5) is satisfied, and if $x, y \in U$ are such that $y \in F(F(\{x\}))$, then, by 3.2, there exists an element $z \in U$ such that $\langle x, z\rangle \in R$ and $\langle z, y\rangle \in R$. Since R is symmetric and is a function, this implies that $\langle z, x\rangle \in R$ and, hence, $x = y$. Thus $F(F(\{x\})) \subseteq \{x\}$ for any $x \in U$, and, consequently, by 1.8(1), $F(F(X)) \subseteq X$ for any $X \in A$. Therefore, by the additivity of F,

$$X \cap F(U) = F(F(X)) \cup [X \cap F(F(X)^-)].$$

Since (i_3) is satisfied, it follows that (i_5) holds.

THEOREM 3.6. *Two functions F and G on* A *to* A *are conjugate if, and only if, there exists a relation $R \subseteq U^2$ such that*

$$F = R^* \text{ and } G = (R^{-1})^*.$$

Proof. If F and G are conjugate, then F and G are normal and completely additive by 1.14. It follows by 3.3 that there exist relations $R, S \subseteq U^2$ such that $F = R^*$ and $G = S^*$. Consider two elements $x, y \in U$ with $\langle x, y\rangle \in R$. Then, by 3.2, $y \in F(\{x\})$ or, equivalently, $F(\{x\}) \cap \{y\} \neq \Lambda$. By 1.11 this gives $G(\{y\}) \cap \{x\} \neq \Lambda$, so that $x \in G(\{y\})$ and, consequently, $\langle y, x\rangle \in S$. We thus have $R^{-1} \subseteq S$. Similarly $S^{-1} \subseteq R$, and we conclude that $S = R^{-1}$.

Assume that $F = R^*$ and $G = (R^{-1})^*$. Consider two sets $X, Y \subseteq U$ with $F(X) \cap Y = \Lambda$. For any elements $x \in X$ and $y \in Y$ we then have $\langle x, y\rangle \notin R$ and, hence, $\langle y, x\rangle \notin R^{-1}$. It follows that $G(Y) \cap X = \Lambda$. Similarly, if $X, Y \subseteq U$ are such that $G(Y) \cap X = \Lambda$, then $F(X) \cap Y = \Lambda$. Hence F and G are conjugate by 1.11.

THEOREM 3.7. *Theorem 3.5 remains valid if we replace there '*A*' by '*B*,' 'completely additive' by 'additive' and 'R^*' by 'R^*/B.' Similarly Theorem 3.6 remains valid if we replace there '*A*' by '*B*,' 'R^*' by 'R^*/B,' and '$(R^{-1})^*$' by '$(R^{-1})^*/\mathsf{B}$.'*

Proof. Suppose F is a normal and additive function on B to B. If F satisfies one or more of the conditions 3.5 (i_1)-(i_6) with 'A' replaced by 'B,' then, by 2.18, the same condition will be satisfied if we replace 'F' by 'F^+' but leave 'A' unchanged. Since, by 2.3 and 2.4, F^+ is normal and completely additive, there exists, by 3.5, a relation $R \subseteq U^2$ which satisfies the corresponding conditions 3.5(ii_1)-(ii_6) and for which $R^* = F^+$ and, hence, $R^*/\mathsf{B} = F$.

If, conversely, there exists a relation $R \subseteq U^2$ such that $F = R^*/\mathsf{B}$, then F is normal and additive by 3.3. If, in addition, R satisfies one or more of the conditions 3.5(ii$_1$)-(ii$_6$), then, by 3.3 and 3.5, the corresponding conditions 3.5(i$_1$)-(i$_6$) will be valid if we replace there 'F' by 'R^*.' Consequently the same conditions will hold if we leave F unchanged but replace 'A' by 'B.' The proof of the second part of the theorem is similar.

DEFINITION 3.8. *By the complex algebra of an algebra in the wider sense*

$$\mathfrak{U} = \langle U, R_0, R_1, \cdots, R_\xi, \cdots \rangle$$

we mean the algebra

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, U, R^*_0, R^*_1, \cdots, R^*_\xi, \cdots \rangle$$

where A *is the family of all subsets of* U.

THEOREM 3.9. *The complex algebra of any algebra in the wider sense is a normal, complete, and atomistic Boolean algebra with operators.*

Conversely, every normal, complete, and atomistic Boolean algebra with operators is isomorphic to the complex algebra of some algebra in the wider sense.

Proof. The first part follows from 2.13, 3.3, and 3.8. Suppose

$$\mathfrak{A} = \langle A, +, 0, \cdot; 1, f_0, f_1, \cdots, f_\xi, \cdots \rangle$$

is a normal, complete, and atomistic Boolean algebra with operators. Let the rank of f_ξ be m_ξ. Then $\mathfrak{A}$ is isomorphic to an algebra

$$\mathfrak{A}' = \langle \mathsf{A}, \cup, \Lambda, \cap, U, F_0, F_1, \cdots, F_\xi, \cdots \rangle$$

where U is some set, A is the family of all subsets of U, and F_ξ is a normal and completely additive function on A^{m_ξ} to A. By 3.33 there exist relations $R_0, R_1, \cdots, R_\xi, \cdots$ such that $R_\xi \subseteq U^{m_\xi+1}$ and $F_\xi = R^*_\xi$. Hence, by 3.8, $\mathfrak{A}'$ is the complex algebra of the algebra in the wider sense

$$\mathfrak{U} = \langle U, R_0, R_1, \cdots, R_\xi, \cdots \rangle.$$

THEOREM 3.10 (REPRESENTATION THEOREM). *Every normal Boolean algebra with operators is isomorphic to a regular subalgebra of the complex algebra of some algebra in the wider sense.*

Proof. If $\mathfrak{B}$ is a normal Boolean algebra with operators, then there exists, by 2.15, a complete and atomistic Boolean algebra with operators $\mathfrak{A}$

which is a perfect extension of $\mathfrak{B}$. By 2.13 and 2.14, $\mathfrak{A}$ is also normal, and the conclusion follows by 3.9.

We now turn to the applications of our results to closure algebras and cylindric algebras.

DEFINITION 3.11. *An algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, C\rangle$$

(where $+$ and $\cdot$ are operations on A^2 to A, C is an operation on A to A, and 0 and 1 are elements of A) is called a closure algebra if it satisfies the following conditions:

(i) *$\langle A, +, 0, \cdot, 1\rangle$ is a Boolean algebra.*

(ii) *C is an additive function on A to A.*

(iii) *$C(C(x)) = C(x)$ for any $x \in A$.*

(iv) *$x + C(x) = C(x)$ for any $x \in A$.*

(v) *$C(0) = 0$.*

Since each of the conditions (iii)-(v) has the form of an equation, we can apply Theorem 2.15 to closure algebras. This gives

THEOREM 3.12. *For any closure algebra $\mathfrak{B}$ there exists a complete and atomistic closure algebra $\mathfrak{A}$ which is a perfect extension of $\mathfrak{B}$.*

Proof. By 2.15, 2.18, and 3.11.

Before applying Theorem 3.10 to closure algebras, we ask when the complex algebra of a given algebra in the wider sense is a closure algebra. The answer to this question is given in the following

THEOREM 3.13. *Suppose U is an arbitrary set, A is the family of all subsets of U, and $R \subseteq U^2$. Then the algebra*

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, U, R^*\rangle$$

is a closure algebra if, and only if, R is reflexive and transitive, and the domain of R is U.[9]

---

[9] This theorem is essentially known from the literature; compare Birkhoff [1], Theorem 1.

Proof. Since, by 3. 3, R^* is normal and completely additive, we see by 3. 11 that $\mathfrak{A}$ is a closure algebra if, and only if,

$$X \subseteq R^*(X) \text{ and } R^*(R^*(X)) \subseteq R^*(X) \text{ for any } X \in \mathsf{A}.$$

Hence the conclusion follows from 3. 3 and 3. 5.

THEOREM 3. 14. *An algebra*

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1, C \rangle$$

is a closure algebra if, and only if, $\mathfrak{B}$ *is isomorphic to an algebraic system*

$$\mathfrak{B}' = \langle \mathsf{B}, \cup, \Lambda, \cap, U, R^* \rangle$$

where B *is a regular set-field whose universal set is* U, *and* R *is a reflexive and transitive relation whose domain is* U *and which satisfies the condition:* $R^*(X) \in \mathsf{B}$ *for every* $X \in \mathsf{B}$ (*i. e.,* R^*/B *is an operation on* B *to* B).

Proof. Assume that $\mathfrak{B}$ is a closure algebra. By 3. 12, there is a complete and atomistic closure algebra $\mathfrak{A}$ which is a perfect extension of $\mathfrak{B}$. By 3. 8, 3. 9, 3. 11, and 3. 13, $\mathfrak{A}$ is isomorphic to an algebra

$$\mathfrak{A}' = \langle \mathsf{A}, \cup, \Lambda, \cap, U, R^* \rangle$$

where U is a certain set, A is the family of all subsets of U, and R is a reflexive and transitive relation whose domain is U. By 2. 14, $\mathfrak{B}$ is a regular subalgebra of $\mathfrak{A}$, and hence it is isomorphic to a regular subalgebra

$$\mathfrak{B}' = \langle \mathsf{B}, \cup, \Lambda, \cap, U, R^* \rangle$$

of $\mathfrak{A}'$. Since, by 3. 11, $\mathfrak{B}$ is a Boolean algebra supplemented by a unary operation, B is a regular set-field, and R^*/B is an operation on B to B.

If, conversely, $\mathfrak{B}$ is isomorphic to an algebra $\mathfrak{B}'$ with the properties listed in the theorem, then, by 3. 13, $\mathfrak{B}$ is isomorphic to a subalgebra of a closure algebra and hence, by 3. 11, it is itself a closure algebra.

The theorem just proved clearly remains valid if we omit in it the term 'regular.' An analogous remark applies to Theorem 3. 18 below.

By a representation theorem for closure algebras which is known from the literature, every closure algebra $\mathfrak{B}$ is isomorphic to an algebra

$$\mathfrak{B}' = \langle \mathsf{B}, \cup, \Lambda, \cap, U, C \rangle$$

where U is a topological space with the closure operation C, and B is a set-field whose universal set is U and which satisfies the condition: $C(X) \varepsilon \mathfrak{B}$ for every $X \varepsilon \mathsf{B}$; in addition, in this topological space we have $C(X) = X$ for every set $X \subseteq U$ consisting of just one point.[10] On the other hand, Theorem 3.14 implies that every closure algebra $\mathfrak{B}$ is isomorphic to an algebra $\mathfrak{B}'$ as described above, in which, however, we do not necessarily have $C(X) = X$ for every one-point set $X \subseteq U$, but in which instead the closure operation C is complete additive. The new representation theorem appears to have interesting implications for every topological space whose closure operation is not completely additive; for it shows that the closure algebra of all subsets of such a space can be isomorphically embedded in a topological space with a completely additive closure operation.

DEFINITION 3.15. *An algebra*

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, C_0, C_1 \rangle$$

(where $+$ *and* $\cdot$ *are operations on* A^2 *to* A, C_0 *and* C_1 *are operations on* A *to* A, *and* 0 *and* 1 *are elements of* A*) is called a cylindric algebra if the following conditions are satisfied:*

(i) $\langle A, +, 0, \cdot, 1 \rangle$ *is a Boolean algebra.*

(ii) C_0 *and* C_1 *are selfconjugate functions (i. e., for any elements* $x, y \varepsilon A$, *if* $C_0(x) \cdot y = 0$, *then* $C_0(y) \cdot x = 0$, *and if* $C_1(x) \cdot y = 0$, *then* $C_1(y) \cdot x = 0$).

(iii) $C_0(C_0(x)) = C_0(x)$ *and* $C_1(C_1(x)) = C_1(x)$ *for any* $x \varepsilon A$.

(iv) $C_0(C_1(x)) = 1$ *for any element* $x \varepsilon A$ *with* $x \neq 0$.[11]

THEOREM 3.16. *For any cylindric algebra* $\mathfrak{B}$ *there exists a complete and atomistic cylindric algebra* $\mathfrak{A}$ *which is a perfect extension of* $\mathfrak{B}$.

Proof. Suppose that

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1, C_0, C_1 \rangle$$

is a cylindric algebra. By 1.14, 2.13, and 3.15, $\mathfrak{B}$ is a Boolean algebra with operators. Hence, by 2.15, there exists a complete and atomistic Boolean algebra with operators

---

[10] Cf. McKinsey-Tarski [1], Theorem 2.6.

[11] A slightly different, though equivalent, definition of these algebras is given in Chin-Tarski [2] (where, however, the term "cylindric algebra" is not used).

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, f_0, f_1 \rangle$$

which is a perfect extension of $\mathfrak{B}$. Thus $\mathfrak{A}$ satisfies 3.15(i). By 2.2, 2.4, 2.13, and 2.14 we have $f_0 = C^+{}_0$ and $f_1 = C^+{}_1$; hence, by 2.10 and 2.12, $\mathfrak{A}$ satisfies 3.15(ii)(iii). Finally, by arguing as in the proof of 2.11(ii), we show that $\mathfrak{A}$ satisfies 3.15(iv). Consequently, $\mathfrak{A}$ is a cylindric algebra.

THEOREM 3.17. *Suppose U is an arbitrary set, A is the family of all subsets of U, $R_0 \subseteq U^2$, and $R_1 \subseteq U^2$. Then the algebra*

$$\mathfrak{A} = \langle \mathsf{A}, \cup, \Lambda, \cap, U, R^*{}_0, R^*{}_1 \rangle$$

is a cylindric algebra if, and only if, the following conditions are satisfied:

(i) *R_0 and R_1 are equivalence relations.*

(ii) $$R_0 \mid R_1 = U^2.$$

Proof. If $\mathfrak{A}$ is a cylindric algebra, then R_0 and R_1 are equivalence relations by 3.3, 3.5, and 3.15(ii)(iii). By 3.2 and 3.15(iv) we have $R_1 | R_0 = U^2$ and, hence, $R_0 | R_1 = (R_1 | R_0)^{-1} = U^2$. Thus (i) and (ii) hold.

If (i) and (ii) are satisfied, then we see by 3.2 that 3.15(iv) holds (with the obvious substitutions) while conditions (ii) and (iii) of 3.15 follow from 3.3 and 3.5 and the fact that R_0 and R_1 are transitive.

THEOREM 3.18. *An algebra*

$$\mathfrak{B} = \langle B, +, 0, \cdot, 1, C_0, C_1 \rangle$$

is a cylindric algebra if, and only if, $\mathfrak{B}$ is isomorphic to an algebraic system

$$\mathfrak{B}' = \langle \mathsf{B}, \cup, \Lambda, \cap, U, R^*{}_0, R^*{}_1 \rangle$$

satisfying the following conditions:

(i) *B is a regular set-field; U is the universal set of B; $R^*{}_0(X) \in \mathsf{B}$ and $R^*{}_1(X) \in \mathsf{B}$ whenever $X \in \mathsf{B}$.*

(ii) *R_0 and R_1 are equivalence relations.*

(iii) $$R_0 \mid R_1 = U^2.$$

Proof. Analogous to that of 3.14, with 3.11, 3.12, and 3.13 replaced by 3.15, 3.16, and 3.17, respectively.

An important modification of Theorem 3.18 is known (which, however, requires an essentially new method of arguing).[12] It can be shown that, after omitting the term "regular" in condition (i), the following condition can be added:

(iv) $R_0 \cap R_1$ is the identity function on U to U.

The result thus obtained admits an interesting geometric interpretation. For, as is easily seen, it implies that the set U and the relations R_0 and R_1 can be chosen in 3.18 in such a way that: (i) U is a set of the form $U = X \times Y$, and hence can be regarded as an abstract two-dimensional analytic space with the axes X and Y; (ii) R_0 is the relation which holds between any two couples $\langle x, y\rangle$ and $\langle x', y'\rangle$ in U if, and only if, $y = y'$, and hence R^*_0 coincides with the so-called operation of cylindrification parallel to the axis X; similarly, (iii) R_1 holds between $\langle x, y\rangle, \langle x', y'\rangle \in U$ if, and only if, $x = x'$, and hence R^*_1 coincides with cylindrification parallel to Y. In this geometric form, the modified Theorem 3.18 motivates the choice of the name given cylindric algebras, and it becomes the natural representation theorem for this class of algebraic systems.

It can readily be shown that every cylindric algebra $\mathfrak{A}$ satisfies the condition

(iv′) $C_0(C_1(x)) = C_1(C_0(x))$ for every $x \in A$.

If we replace (iv) by (iv′) in 3.15, we obtain the definition of a more comprehensive class of algebras, which can be referred to as *generalized cylindric algebras*. Cylindric algebras as originally defined prove to coincide with those generalized cylindric algebras which are simple (in the general algebraic sense). Theorems 3.16-3.18 extend to generalized cylindric algebras; obviously, formulas 3.17(ii) and 3.18(iii) have to be replaced by $R_0 \mid R_1 = R_1 \mid R_0$. Theorem 3.18 can again be improved so as to admit a geometric interpretation. In this interpretation the set U is assumed to coincide, not necessarily with the whole two-dimensional space $X \times Y$, but with a subset of $X \times Y$ (subjected to some additional conditions), while R_0 and R_1 preserve their old meaning.

On the other hand, *projective algebras* (mentioned in the introduction) can be regarded as specialized cylindric algebras. In fact, a projective algebra can be characterized as an algebraic system

$$\mathfrak{A} = \langle A, +, 0, \cdot, 1, C_0, C_1, \circledcirc\rangle$$

---

[12] In fact, the method used in Everett-Ulam [1], Section 4, pp. 85-88, can be applied here.

satisfying the conditions:

(i) $\langle A, +, 0, \cdot, 1, C_0, C_1 \rangle$ is a cylindric algebra,

(ii) Θ is an atom in A,

(iii) $x = C_0(x) \cdot C_1(x)$ whenever $x \,\varepsilon\, A$ and $x \leq C_0(\Theta)$ or $x \leq C_1(\Theta)$.

In terms of C_0, C_1, and Θ two further operations on A to A can be defined:

$$P_0(x) = C_0(x) \cdot C_1(\Theta) \text{ and } P_1(x) = C_1(x) \cdot C_0(\Theta) \text{ for every } x \,\varepsilon\, A.$$

Conversely, C_0, C_1, and Θ can be re-defined in terms of P_0 and P_1; hence, projective algebras can also be characterized as Boolean algebras supplemented by these two new operations. Again Theorems 3.16-3.18 extend, with obvious changes, to projective algebras; and the extended Theorem 3.18 can be improved so as to admit a geometric interpretation (and thus to yield the natural representation theorem for these algebras). In this interpretation U, R_0, and R_1 have the same meaning as in the case of cylindric algebras. Θ is represented as the set consisting of a single couple $\langle x_0, y_0 \rangle$ for some $x_0 \,\varepsilon\, X$ and $y_0 \,\varepsilon\, Y$; this couple can be regarded as the origin of the coordinate system in the space $U = X \times Y$. Finally, P_0 and P_1 are represented as the operations of projection parallel to the axes X and Y.[13]

Brown University, Providence, Rhode Island,
University of California, Berkeley, California.

---

[13] Our definition of projective algebras is essentially equivalent to the one given in Everett-Ulam [1] where, however, P_0 and P_1 (instead of C_0, C_1 and Θ) are treated as fundamental operations. A definition of projective algebras in terms of C_0, C_1, and Θ can be found in Chin-Tarski [2]; however, Mr. F. B. Thompson has pointed out to us that in this definition the postulates involving Θ are too weak and should be modified in the way indicated in the text. The improved Theorem 3.18 in its geometric form, when extended to projective algebras, essentially coincides with the representation theorem which has been established in Everett-Ulam [1] for complete atomistic projective algebras and generalized in McKinsey [1] to all projective algebras.

ORTHOGONAL AND UNITARY GROUPS OVER THE RATIONAL FIELD.*

By JEAN DIEUDONNÉ.

1. Let E be an n-dimensional space over the field $\boldsymbol{Q}$ of rational numbers; let f be a non-degenerate symmetric bilinear form defined in E, and $g(x) = f(x, x)$ the corresponding quadratic form. In a recent paper [2],[1] I have proved that for $n \geqq 7$, the commutator subgroup $P\Omega_n(\boldsymbol{Q}, f)$ of the projective orthogonal group $\boldsymbol{PO}_n(\boldsymbol{Q}, f)$ is always *simple*, even when the index ν of f is 0. By a modification of the argument used in the proof of that theorem, I shall now show that the result is still true for $n = 6$. Using the well-known isomorphism between orthogonal groups on 6 variables and unitary groups on 4 variables, this enables me to prove that if K_1 is any quadratic extension of $\boldsymbol{Q}$, E_1 an n-dimensional space over K_1, f_1 a non-degenerate hermitian form over E_1, the commutator subgroup of the projective unitary group $\boldsymbol{PU}^{+}_n(K_1, f_1)$ is *simple* for $n \geqq 5$, even when the index of f_1 is 0.

2. The proof given in [2] uses essentially a theorem of Meyer to the effect that a non-degenerate quadratic form in at least 4 variables over the rational field can take *any* rational value if it is indefinite, any positive (resp. negative) rational value if it is positive definite (resp. negative definite). In applying that theorem, I have to use repeatedly the assumption $n \geqq 6$; however, the assumption $n \geqq 7$ is used essentially only twice, in steps 3° and 4° of the proof. I proceed to show that, by a modification of the argument, these steps too can be carried out for $n = 6$. I keep the notation of [2]. G being a normal subgroup of $\Omega_n(\boldsymbol{Q}, f)$ not contained in the center, step 1° of the proof shows that there exists in G a plane rotation u distinct from the identity. If u^2 is the identity, u is an $(n-2, 2)$-involution in Ω_n, and therefore [2, n° 10] a special involution; in that case, step 3° is proved. If u is not an involution, let P be its plane, a and b two orthogonal vectors in P, and let $u(a) = \lambda a + \mu b$; let $\alpha = g(a)$, $\beta = g(b)$ (which are positive rational numbers). If E' is the space over the real field $\boldsymbol{R}$ obtained from E by extension of the field of scalars, u is extended to a plane rotation in E', of an

---

* Received May 14, 1951.

[1] Numbers in brackets refer to the bibliography at the end of the paper.

angle θ such that $\cos\theta = \lambda$, $\sin\theta = \mu(\beta/\alpha)^{\frac{1}{2}}$; replacing if necessary u by a power u^m of u, it can always be supposed that $\pi/2 < \theta < \pi$ [2, n° 3]. Consider the vector $c = \frac{1}{2}(a + w(a)) = \frac{1}{2}((1+\lambda)a + \mu b)$; it is easily verified that $u = s's$, where s is the symmetry with respect to the hyperplane orthogonal to a, and s' the symmetry with respect to the hyperplane orthogonal to c. We are going to show that there exists a vector x such that $g(x) = g(c)$, $f(a,x) = f(a,c)$ and such that x and c are orthogonal, that is, $f(x,c) = 0$. To that purpose, let y be a vector in the hyperplane orthogonal to a, and write $x = \frac{1}{2}((1+\lambda)a + \mu y)$; this already insures that $f(a,c) = f(a,x)$. The relation $f(x,c) = 0$ then becomes $\alpha(1+\lambda)^2 + \mu^2 f(b,y) = 0$; write $y = \rho b + z$, where $z \varepsilon P^*$ (orthogonal to both a and b); this yields

$$\rho = -(\alpha/\beta)(1+\lambda)^2/\mu^2 = -(1+\cos\theta)^2/\sin^2\theta = -\operatorname{cotg}^4\theta/2$$

Finally, the relation $g(x) = g(c)$ is equivalent to $g(y) = g(b)$, which in turn gives $g(z) = (1-\rho^2)g(b)$. Now, as $\pi/4 < \theta/2 < \pi/2$, one has $|\operatorname{cotg}\theta/2| < 1$, and therefore $\rho^2 < 1$; as P^* is at least 4-dimensional, Meyer's theorem shows that a vector z satisfying the preceding relation can be found in P^*.

It follows from Witt's theorem that there exists an orthogonal transformation $v \varepsilon \boldsymbol{O}_n$ such that $v(a) = a$ and $v(c) = x$; the argument in [2, n° 5] (which only uses the assumption $n \geqq 6$) proves then that there exists a transformation $w \varepsilon \boldsymbol{O}_n$ leaving invariant a and x, and such that $v_1 = wv$ is in the commutator subgroup Ω_n. The transformation $u_1 = wv_1u^{-1}v_1^{-1}$ belongs therefore to G; but $v_1u^{-1}v_1^{-1} = ss''$, where $s'' = v_1s'v_1^{-1}$ is the symmetry with respect to the hyperplane orthogonal to x. Therefore $u_1 = s'sss'' = s's''$ is a special involution in G (since x and c are orthogonal), and this proves point 3°.

3. We now prove point 4°, which states that G contains *every* special involution. To achieve this, we notice that, in the preceding construction, we had $g(c) = \frac{1}{4}((1+\lambda)^2\alpha + \mu^2\beta) = \frac{1}{2}\alpha(1+\lambda) = \frac{1}{2}f(a, a+u(a))$. All we have to do is therefore to find a plane rotation u in G and a vector z in its plane such that $f(z,u(z)) < 0$ and that $\frac{1}{2}f(z, z+u(z))$ is equal to a given positive rational number γ_0 multiplied by a square. We start from a special involution w in G, of plane R; let b be a vector in R, and consider all vectors $z \varepsilon E$ of the form $z = a + b$, where $a \varepsilon R^*$; we write $\alpha = f(a,a)$, $\beta = f(b,b)$. For any such vector z, there exists in R^* (which is at least 4-dimensional) a vector y such that $f(y,y) = f(z,z)$, by Meyer's theorem. Let s, s' be the symmetries with respect to the hyperplane orthogonal to z and y; they are conjugate in $\boldsymbol{O}_n$ by Witt's theorem, hence $v = s's$ is in Ω_n.

It follows that $u = wv^{-1}w^{-1}v$ is in G; but as $w(y) = y$, $wv^{-1}w^{-1} = s''s'$, where s'' is the symmetry with respect to the hyperplane orthogonal to $w(z) = a - b$; therefore $u = s''s$ is a plane rotation in G, z is in its plane, and $u(z) = -s''(z) = -z + 2f(z, w(z))w(z)/g(z)$. This formula shows that the conditions we have to satisfy are $(\alpha - \beta)^2/(\alpha + \beta) = \gamma_0\xi^2$, where $\xi \in \boldsymbol{Q}$, and $\alpha + \beta > 2^{\frac{1}{2}}(\alpha - \beta)$. Now it is easily verified that the first requirement is satisfied by taking $\xi = t - 2\beta/(\gamma_0 t)$, $\alpha + \beta = \gamma_0 t^2$, $\alpha - \beta = \gamma_0\xi t$, where t is any positive rational number; the condition $\alpha > 0$ reduces to $t^2 > \beta/\gamma_0$, and the condition $\alpha + \beta > 2^{\frac{1}{2}}(\alpha - \beta)$ to $t^2 < 2(2^{\frac{1}{2}})/(2^{\frac{1}{2}} - 1) \cdot \beta/\gamma_0$. These conditions are therefore compatible, and with a value of t which satisfies them, the rotation u will satisfy our requirements, provided there is a vector $a \in R^*$ such that $f(a, a) = \alpha$; but since $\alpha > 0$ and R^* is at least 4-dimensional, the existence of a is proved by Meyer's theorem.

We have thus proved the existence in G of a special involution conjugate (within $\boldsymbol{O}_n$) of any special involution. The end of the proof [2, end of n° 5 and n° 6] only makes use of the relation $n \geqq 6$, and we see therefore that *for any form f of index 0 in E, the group $P\Omega_6(\boldsymbol{Q}, f)$ is simple.*

Of course, the remarks made in [2, n° 7] still apply here, and $P\Omega_6(K, f)$ is simple for any subfield K of $\boldsymbol{R}$ in which Meyer's theorem is valid. For the field $\boldsymbol{Q}$, the theorem is known to be false for $n = 3$ and $n = 4$ [1, p. 38]; the only case which remains open in that case is $n = 5$; an entirely new approach seems to be needed for that case, since the assumption $n \geqq 6$ is essential to all but the first two steps of our proof.

4. We now turn to the *unitary groups* $\boldsymbol{U}_n(K_1, f_1)$ over quadratic number fields K_1. For an hermitian non-degenerate form f_1 in an n-dimensional space E_1 over the field K_1, there is a basis $(e_i)_{1 \leqq i \leqq n}$ of E_1 such that $f_1(z, z) = \sum_{k=1}^{n} \alpha_k \zeta_k \bar{\zeta}_k$, where the α_k are *rational* numbers; if $K_1 = \boldsymbol{Q}(\omega)$, where $\omega^2 = \rho$ is rational, and $\zeta_k = \xi_k + \omega\eta_k$, ξ_k and η_k rational, one has $f_1(z, z) = \sum_{k=1}^{n} \alpha_k(\xi_k{}^2 - \rho\eta_k{}^2)$. This, together with Meyer's theorem, shows at once that if $n \geqq 3$, and either $\rho > 0$ or the α_k have not all the same sign, then the form f_1 has an index $\geqq 1$; I have shown that in such a case, the projective unitary group $\boldsymbol{PU}^+{}_n(K_1, f_1)$ is simple [1, p. 70]. The only remaining case is therefore that in which K_1 is an imaginary quadratic field and the form f_1 is positive definite.

We shall always suppose, in what follows, that $n \geqq 4$ and the index ν of f_1 is 0. We begin by proving some preliminary lemmas.

LEMMA 1. *The commutator subgroup Υ_n of $\boldsymbol{U}_n(K_1, f_1)$ is identical with the commutator subgroup Υ^+_n of $\boldsymbol{U}^+_n(K_1, f_1)$.*

One has obviously $\Upsilon^+_n \subset \Upsilon_n$, and the only thing to prove is that every element in Υ_n belongs to Υ^+_n. Let us call *quasi-symmetries* the transformations in $\boldsymbol{U}_n$ which leave invariant all elements of a hyperplane in E_1; if $(e_i)_{1 \leqq i \leqq n}$ is an orthogonal basis of E_1 such that the e_i of index $\leqq n-1$ form a basis of a hyperplane H, e_n being orthogonal to H, any transformation w leaving invariant every element of H and belonging to $\boldsymbol{U}_n$, is such that $w(e_i) = e_i$ for $1 \leqq i \leqq n-1$ and $w(e_n) = \delta e_n$ with $\delta\bar{\delta} = 1$, and conversely. It follows immediately that every transformation in $\boldsymbol{U}_n$ can be written sv where $v \,\varepsilon\, \boldsymbol{U}^+_n$ and s is a quasi-symmetry of fixed hyperplane H; we are going to see that it is enough to prove that every commutator $svs^{-1}v^{-1}$ of a quasi-symmetry of hyperplane H and a transformation $v \,\varepsilon\, \boldsymbol{U}^+_n$ belongs to Υ^+_n. Indeed, let Γ be the subgroup of Υ_n generated by Υ^+_n and these commutators; Γ is a normal subgroup in $\boldsymbol{U}_n$, for one has, for a quasi-symmetry s_0 of hyperplane H,

$$s_0(svs^{-1}v^{-1})s_0^{-1} = (s_0s)v(s_0s)^{-1}v^{-1} \cdot vs_0v^{-1}s_0^{-1}$$

and for any $v_0 \,\varepsilon\, \boldsymbol{U}^+_n$

$$v_0(svs^{-1}v^{-1})v_0^{-1} = sus^{-1}u^{-1} \cdot s^{-1}v_0sv_0^{-1}$$

with $u = s^{-1}v_0sv \,\varepsilon\, \boldsymbol{U}^+_n$. In the quotient group $\boldsymbol{U}_n/\Gamma$, the classes of quasi-symmetries (of hyperplane H) commute therefore with all classes of elements of $\boldsymbol{U}^+_n$ but these latter commute among themselves, and so do the quasi-symmetries of hyperplane H; therefore, as $\boldsymbol{U}_n/\Gamma$ is generated by all these classes, it is an abelian group, and Γ contains Υ_n, which proves that it is identical to Υ_n. We thus have only to prove that $svs^{-1}v^{-1} \,\varepsilon\, \Upsilon^+_n$, or in other words, if s_1 and s_2 are two conjugate quasi-symmetries in $\boldsymbol{U}_n$, that $s_1s_2^{-1} \,\varepsilon\, \Upsilon^+_n$. Suppose s_1 (resp. s_2) is a quasi-symmetry of hyperplane H_1 (resp. H_2); by assumption, there is a vector a_1 orthogonal to H_1 and a vector a_2 orthogonal to H_2 such that $f_1(a_1, a_1) = f_1(a_2, a_2)$, and $s_1(a_1) = \alpha a_1$, $s_2(a_2) = \alpha a_2$, with $\alpha\bar{\alpha} = 1$. The intersection $H_1 \cap H_2$ is at least 2-dimensional over K_1, hence at least 4-dimnsional over $\boldsymbol{Q}$; by Meyer's theorem, there exists in $H_1 \cap H_2$ a vector c such that $f_1(c, c) = f_1(a_1, a_1) = f_1(a_2, a_2)$. Let s_0 be the quasi-symmetry of hyperplane H_0 orthogonal to c, and such that $s_0(c) = \alpha c$; one can write $s_1s_2^{-1} = s_1s_0s_0^{-1}s_2^{-1} = s_1s_0(s_2s_0)^{-1}$. By Witt's theorem, there is a transformation $w \,\varepsilon\, \boldsymbol{U}_n$ such that $w(a_1) = a_2$, $w(c) = c$; if w has a determinant $\delta \neq 1$, multiplying w by a quasi-symmetry of determinants δ^{-1}, whose hyperplane contains a_2 and c, gives a transformation $v \,\varepsilon\, \boldsymbol{U}^+_n$ such that

$v(a_1) = a_2$, $v(c) = c$; therefore one has $s_2 s_0 = v(s_1 s_0) v^{-1}$; as $s_1 s_0$ belongs to $\boldsymbol{U}^+_n$, this ends the proof of Lemma 1.

5. LEMMA 2. *Let P and P' be two planes in E_1 such that the restrictions of f_1 to P and P' are equivalent. Then, if $n \geqq 5$, there is a transformation v in the commutator subgroup Υ_n such that $v(P) = P'$.*

This is of course a reinforcement of Witt's theorem, which merely asserts the existence of v in $\boldsymbol{U}_n$ having the required property. Let a, b be an orthogonal basis of P, and a', b' an orthogonal basis of P' such that $f_1(a', a') = f_1(a, a)$ and $f_1(b', b') = f_1(b, b)$. Let us prove first there is a $w_1 \varepsilon \Upsilon_n$ such that $w_1(a) = a'$. Indeed, the subspace M orthogonal to a and a' is at least 2-dimensional over K_1, hence at least 4-dimensional over $\boldsymbol{Q}$; by Meyer's theorem there is a vector $c \varepsilon M$ such that $f_1(c, c) = f_1(a, a) = f_1(a', a')$. As a and c are orthogonal, there is a symmetry s such that $s(a) = c$, namely the symmetry with respect to the hyperplane orthogonal to $d = c - a$; similarly, the symmetry s' with respect to the hyperplane orthogonal to $d' = c - a'$ is such that $s'(c) = a'$. Now, as $f_1(d, d) = f_1(d', d')$, it follows from Witt's theorem that s and s' are conjugate in $\boldsymbol{U}_n$; therefore $w_1 = s's = s's^{-1} \varepsilon \Upsilon_n$, and one has $w_1(a) = a'$.

We can now suppose that $a = a'$, and show the existence of a transformation $w_2 \varepsilon \Upsilon_n$ such that $w_2(a) = a$ and $w_2(b) = b'$. If H is the hyperplane orthogonal to a, H is at least 4-dimensional over K_1, and contains b and b'; the preceding argument may then be repeated in H instead of E_1, and this proves the lemma.

6. LEMMA 3. *For $n \geqq 2$ (and $\nu = 0$), the group $\boldsymbol{U}^+_n(K_1, f_1)$ is generated by plane rotations.*[2]

This is a simple consequence of Witt's theorem. Let us apply induction on n, the result being trivial for $n = 2$. Let u be any transformation in $\boldsymbol{U}^+_n$, x any vector $\neq 0$; if $u(x) = x$, the hyperplane H orthogonal to x is invariant by u, hence the restriction of u to H is a product of plane rotations by the inductive hypothesis, and so is u itself. If $u(x) \neq x$, let P be a plane containing x and $u(x)$; as $f_1(x, x) = f_1(u(x), u(x))$, there is a transformation $v \varepsilon \mathbf{U}_n$, leaving invariant every element of P^*, and such that $v(x) = u(x)$; if v has a determinant $\delta \neq 1$, let y be a vector in P orthogonal to x, and consider the quasi-symmetry s of hyperplane orthogonal to y and such that $s(y) = \delta^{-1}y$; $w = vs$ is then a plane rotation of plane P, such that $w(x) = u(x)$. Applying the inductive hypothesis to $w^{-1}u$ gives the result.

---

[2] The statement made (without proof) in [1, p. 66] to the effect that every transformation in $U^{\pm}_n(K_1, f_1)$ is a product of at most n symmetries, *is not correct*. For instance, if $n = 2$, and α is an element in K such that $\alpha^2 = -1$, $\alpha\bar{\alpha} = 1$, it is easy to verify that the transformation $x \to \alpha x$ cannot be the product of 2 symmetries. I do not even know if in general (when $\nu = 0$) the group $U^{\pm}_2$ is generated by symmetries. This of course is equivalent to the following question: when x and y are two vectors such that $f_1(x, x) = f_1(y, y)$, does there exist a product of symmetries sending x into y? Certainly one symmetry alone cannot do it, for the two vectors $x + y$ and $x - y$ are not orthogonal in general (in contrast to what happens for the orthogonal groups in similar circumstances).

LEMMA 4. *For $n \geqq 5$, the commutator subgroup $\Upsilon_n(K_1, f_1)$ is generated by products of the type $v_1 v_2^{-1}$, where v_1 and v_2 are conjugate plane rotations, whose planes are both in a 4-dimensional subspace of E_1 such that the restriction of f_1 to that subspace has a square discriminant.*

It is easy to show that Υ_n is generated by products $u_1 u_2^{-1}$ where u_1 and u_2 are *arbitrary* conjugate plane rotations. Indeed, in any commutator $uwu^{-1}w^{-1}$ of elements of $\mathbf{U}^+{}_n$, u is a product of a certain number p of plane rotations (Lemma 3); the result being obvious for $p = 1$, one has only to argue inductively on p (see [1], p. 23); if $u = st$, where s is a plane rotation and t a product of $p - 1$ plane rotations, one has

$$uwu^{-1}w^{-1} = s(twt^{-1}w^{-1})s^{-1} \cdot sws^{-1}w^{-1}$$

which proves our assertion.

Let now u_1 and u_2 be conjugate plane rotations, and let M be a 5-dimensional subspace of E_1 containing the planes P_1, P_2 of u_1 and u_2. Let us remark that, given any positive rational numbers α, β, and any 3-dimensional subspace N of E_1, one can always find in N two orthogonal vectors a, b such that $f_1(a, a) = \alpha$, $f_1(b, b) = \beta$, by two successive applications of Meyer's theorem. Therefore, we can first find in $P^*{}_1 \cap M$ a plane Q_1 such that the restriction of f_1 to $N_1 = P_1 + Q_1$ has a square discriminant, and similarly we can find a plane $Q_2 \subset P^*{}_2 \cap M$ such that the restriction of f_1 to $N_2 = P_2 + Q_2$ has a square discriminant. Now, as N_1 and N_2 are both in M and are 4-dimensional, their intersection N is at least 3-dimensional; therefore, there exists a plane $P_3 \subset N$ such that the restriction of f_1 to P_3 is equivalent to its restrictions to P_1 and P_2. There is therefore, by Witt's theorem, a plane rotation u_3 having P_3 as its plane, and conjugate to both u_1 and u_2 in $\mathbf{U}_n$; as $u_1 u_2^{-1} = (u_1 u_3^{-1})(u_3 u_2^{-1})$, our lemma is proved.

7. LEMMA 5. *When $n = 4$ and f_1 has a square discriminant, the group $P\Upsilon_4(K_1, f_1)$ is simple, and Υ_4 is generated by the squares of plane rotations.*

We can always multiply f_1 by a positive rational number α, and choose a basis in E_1 such that $g_1 = \alpha f_1$ is given by the expression

$$g_1(z, z) = \zeta_1\bar{\zeta}_1 + a_2a_3\zeta_2\bar{\zeta}_2 + a_3a_1\zeta_3\bar{\zeta}_3 + a_1a_2\zeta_4\bar{\zeta}_4,$$

where a_1, a_2, a_3 are positive rational numbers. The well-known isomorphism between orthogonal groups in 6 variables and unitary groups in 4 variables ([4, p. 24] and [3, n° 33]) is such that g_1 corresponds to the positive definite symmetric bilinear form $g(x, y)$ in 6 variables over $\boldsymbol{Q}$ defined by

$$g(x, x) = a_1\xi_1^2 - \rho a_1\xi_2^2 + a_2\xi_3^2 - \rho a_2\xi_4^2 + a_3\xi_5^2 - \rho a_3\xi_6^2$$

in the following way; there exists an homomorphism of $\boldsymbol{U}^+_4(K_1, g_1)$ onto a subgroup Γ of $\boldsymbol{O}^+_6(\boldsymbol{Q}, g)$ such that $\boldsymbol{O}^+_6/\Gamma$ is abelian (which means that $\Omega_6 \subset \Gamma$), the kernel Θ of that homomorphism consisting of the identity and the transformation $z \rightarrow -z$. The subgroup Δ of $\boldsymbol{U}^+_4(K_1, g_1)$ generated by the squares of the elements of $\boldsymbol{U}^+_4$ (or the group Δ/Θ if $\Theta \subset \Delta$) is therefore isomorphic to the group generated by the squares of the elements of Γ, hence to the commutator subgroup $\Omega_6(\boldsymbol{Q}, g)$, since that group contains the squares of the elements of $\boldsymbol{O}^+_6$ [1, pp. 23-24] and is simple over its center (n° 3). Therefore Δ is also simple over its center, and as it contains Υ_4 (since $\boldsymbol{U}^+_4/\Delta$ is abelian), Υ_4 is simple over its center and identical to Δ. Moreover, Υ_4 contains the squares of all plane rotations, hence is generated by them because of the simplicity of $P\Upsilon_4$.

8. We can now proceed to the proof of our main

THEOREM. *For any positive definite hermitian form f_1 in $n \geqq 5$ variables over an imaginary quadratic extension K_1 of the rational field, the factor group $P\Upsilon_n(K_1, f_1)$ of $\Upsilon_n(K_1, f_1)$ by its center $\Upsilon_n \cap Z_n$ is simple.*

We first remark that the square of every plane rotation belongs to Υ_n. Indeed, let P be the plane of a rotation v; by Meyer's theorem, there is a 4-dimensional subspace L containing P and such that the restriction f_0 of f_1 to L has a square discriminant. But the subgroup of $\boldsymbol{U}^+_n(K_1, f_1)$ leaving every element of L^* invariant is isomorphic to (and can be identified with) the group $\boldsymbol{U}^+_4(K_1, f_0)$; hence Lemma 5 shows that the square of v belongs to commutator subgroup of $\boldsymbol{U}^+_4(K_1, f_0)$, and a fortiori to Υ_n.

Let now G be a normal subgroup in Υ_n, not contained in the center. Our first step consists in showing that if $u \in G$ is not in the center, there exists a plane $P \subset E_1$ such that $u(P) \neq P$ and $M = P + u(P)$ is 3-dimensional. Owing to the initial remark, this can be done by exactly the same argument as in [2, n° 2], which we therefore suppress here (the only difference consisting in replacing the bilinear form $f(x, y)$ by $\frac{1}{2}(f_1(x, y) + f_1(y, x))$ in the calculation).

As the space M^* is at least 2-dimensional, there exists, by Meyer's theorem, a vector $c \varepsilon M^*$ such that the restriction f_2 of f_1 to the 4-dimensional subspace F generated by M and c has a square discriminant. Let v_1 be a plane rotation of plane P, such that $v = v_1^2$ is not the identity; v belongs to Υ_n and to the commutator subgroup of $\boldsymbol{U}^+_4(K_1, f_2)$. It follows that $u_1 = v^{-1}uvu^{-1}$ belongs to G, is not in the center, and leaves invariant every element of M^*; moreover, as uvu^{-1} is the square of a plane rotation of plane $u(P) \subset F$, it belongs also to the commutator subgroup of $\boldsymbol{U}^+_4(K_1, f_2)$, and so does u_1. Lemma 5 then proves that G contains every element in that commutator subgroup, and in particular G contains *every square of a plane rotation* whose plane is contained in F. But if w is the square of *any* plane rotation of plane Q, there exists a transformation $s \varepsilon \Upsilon_n$ such that sws^{-1} has its plane contained in F: for Meyer's theorem proves the existence in F of a plane Q' such that the restriction of f_1 to Q' is equivalent to the restriction of f_1 to Q, and then Lemma 2 provides a transformation $s \varepsilon \Upsilon_n$ such that $s(Q) = Q'$. Therefore, G contains the square of *every* plane rotation.

Now let t_1 and t_2 be two conjugate plane rotations whose planes are contained in a 4-dimensional subspace N of E_1 such that the restriction f_3 of f to N has a square discriminant; Lemma 5 shows that the commutator subgroup of $\boldsymbol{U}^+_4(K_1, f_3)$ is simple over its center. Moreover, if w is the square of a plane rotation whose plane is in N, w belongs to the commutator subgroup of $\boldsymbol{U}^+_4(K_1, f_3)$. But we have seen that w belongs to G, hence it follows from Lemma 5 that G contains the commutator subgroup of $\boldsymbol{U}^+_4(K_1, f_3)$, and in particular it contains $t_1t_2^{-1}$ (for the restrictions of t_1 and t_2 to N are conjugate *within the group* $\boldsymbol{U}_4(K_1, f_3)$ by Witt's theorem, the restrictions of f_3 to their planes being equivalent). Lemma 4 proves finally that G is identical to Υ_n, and this completes the proof of the theorem.

As a corollary, it follows from that proof that Υ_n is *generated by the squares of the rotations in* $\boldsymbol{U}^+_n$.

The theorem (and its proof) are of course valid for any subfield of $\boldsymbol{R}$ in which Meyer's theorem is true.

9. When a hermitian form f_1 has an index $\nu \geqq 1$, it is known [1, p. 70] that, for any field K_1 and any dimension $n \geqq 2$ (with the exceptions $n = 2$, $K_1 = \boldsymbol{F}_4$ or $K_1 = \boldsymbol{F}_9$, and $n = 3$, $K_1 = \boldsymbol{F}_4$), the commutator subgroup Υ^+_n of $\boldsymbol{U}^+_n$ is *identical with* $\boldsymbol{U}^+_n$. That this is *not* always the case when the index

$\nu = 0$ can be seen on the following example: consider, over the quadratic field $K_1 = \boldsymbol{Q}(2^{\frac{1}{2}})$, the hermitian form in 2 variables $f_1(z, z) = \zeta_1\bar{\zeta}_1 + 5\zeta_2\bar{\zeta}_2$. The unitary group $\boldsymbol{U}^+{}_2(K_1, f_1)$ is then isomorphic to a subgroup Γ of the orthogonal group $\boldsymbol{O}^+{}_4(\boldsymbol{Q}, f)$ relative to the form $f(x, x) = \xi_1{}^2 + 2\xi_2{}^2 + 5(\xi_3{}^2 + 2\xi_4{}^2)$, the isomorphism being defined in the well-known fashion which, to every unitary matrix

$$\begin{pmatrix} \alpha & 5\beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix}$$

in $\boldsymbol{U}^+{}_2(K_1, f_1)$ (with $\alpha\bar{\alpha} + 5\beta\bar{\beta} = 1$) associates the orthogonal matrix of the corresponding linear substitution on the ξ_k, related to ζ_1 and ζ_2 by $\zeta_1 = \xi_1 + \omega\xi_2$ and $\zeta_2 = \xi_3 + \omega\xi_4$, with $\omega^2 = -2$. Now, to the transformation $z \to -z$, which is in $\boldsymbol{U}^+{}_2(K_1, f_1)$, is thus associated the symmetry $x \to -x$ in Γ; but I have proved [2, n° 11], that that transformation does not belong to the commutator subgroup $\Omega_4(\boldsymbol{Q}, f)$, and a fortiori it cannot belong to the commutator subgroup of Γ; hence $\boldsymbol{U}^+{}_2(K_1, f_1)$ is in that case distinct from its commutator subgroup. It would be interesting to know if, for $n \geqq 5$, $\nu = 0$ and K_1 an imaginary quadratic extension of $\boldsymbol{Q}$, the commutator subgroup Υ_n of $\boldsymbol{U}^+{}_n(K_1, f_1)$ is or is not identical to $\boldsymbol{U}^+{}_n$. I have not been able to find an answer to this question.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] J. Dieudonné, "Sur les groupes classiques," *Actualités Scientifiques et industrielles*, no. 1040, Paris (Hermann), 1948.

[2] ———, "On the orthogonal groups over the rational field," *Annals of Mathematics*, vol. 54 (1951), pp. 85-93.

[3] ———, "Les extensions quadratiques des corps non commutatifs et leurs applications," to be published in *Acta Mathematica*.

[4] B. L. van der Waerden, "Gruppen von linearen Transformationen," *Ergebnisse der Mathematik*, Bd. IV, no. 2, Berlin (Springer), 1935.

ON GEODESIC COORDINATES.*

By PHILIP HARTMAN.**

1. The problem. Let $S: X = X(u, v)$ be a surface of class C''. That is, let $X = (x, y, z)$ be a vector in 3-space, and $X = X(u, v)$ a function of class C'' in a vicinity of $(u, v) = (0, 0)$ such that the vector product (X_u, X_v) is not zero (the subscripts u, v denote partial differentiation). Let the squared element of arc-length on S be

$$(1) \qquad ds^2 = g_{ik}(u, v)\, du^i du^k, \qquad (u^1, u^2) = (u, v),$$

where, without loss of generality, it can be supposed that

$$(2) \qquad g_{11}(0, 0) = 1, \qquad g_{12}(0, 0) = 0, \qquad g_{22}(0, 0) = 1.$$

The functions g_{11}, g_{12}, g_{22}, which are the scalar products $X_u \cdot X_u$, $X_u \cdot X_v$, $X_v \cdot X_v$, respectively, are of class C'.

It has recently been shown [1] that it is possible to introduce geodesic polar coordinates (r, ϕ) of class C' at the point $(0, 0)$. By this is meant that there exist functions $u = u(r, \phi)$, $v = v(r, \phi)$ of class C' on a set $0 \leqq r \leqq$ const., $-\infty < \phi < \infty$, with the properties that, for a fixed ϕ, the arc $u = u(r, \phi)$, $v = v(r, \phi)$ is the unique geodesic satisfying the initial conditions

$$(3) \qquad u(0, \phi) = 0,\ v(0, \phi) = 0 \text{ and } u_r(0, \phi) = \cos\phi,\ v_r(0, \phi) = \sin\phi;$$

r is the arc-length on this geodesic; the Jacobian $\partial(u, v)/\partial(r, \phi)$ does not vanish for $r \neq 0$; finally, the transformation $(u, v) \rightarrow (r\cos\phi, r\sin\phi)$ is one-to-one. It was also shown that the functions $u_r(r, \phi)$, $v_r(r, \phi)$ are of class C'. It was left an open question whether or not the transformation $(u, v) \rightarrow (r, \phi)$ is of class C'' (that is, whether or not the partial derivatives $u_{\phi\phi}$, $v_{\phi\phi}$ exist and are continuous; although it was pointed out that, for a fixed r, the arc $u = u(r, \phi)$, $v = v(r, \phi)$ is of class C'' in terms of its arc-length parameter).

The object of this paper is to give an example showing that the answer to the question can be in the negative. The example S, to be given below, is curious for additional reasons. First, the coefficient functions g_{ik} in (1) are of class C''; cf. (4) and (5) below. Of course, S cannot be of class C'''

---

* Received December 20, 1950.

** John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

(by the remark preceding the statement of Theorem 2 in [1], p. 724), nor can the Gaussian curvature of S be of class C' (by § 6 in [1] and the Lemma, § 12, [4]).

Another curious property of S is that, at the point $(u, v) = (0, 0)$, there exists one choice of geodesic parallel coordinates which is not of class C'', while all other choices are of class C''.

The example S turns out to be a surface of revolution and is only a slight modification of the one given in [1], p. 724, to show that if the functions g_{ik} in the Riemannian metric (1) are of class C' (but (1) does not belong to a surface of class C'', nor does (1) possess a bounded Gaussian curvature; cf. Theorem (III) in [3], p. 133), then the geodesics need not be uniquely determined by initial conditions; cf. also the example in [2], p. 554 and the Remark in [3], p. 138.

2. The example. Let $h(v) > 0$ be defined for small $| v |$ by

$$h(v) = 1 + | v |^{\lambda}, \text{ where } 2 < \lambda < 3; \tag{4}$$

so that

$$h(0) = 1 \text{ and } h_v(0) = 0. \tag{5}$$

It is clear that $h(v)$ is of class C'' and, in fact, its second derivative satisfies a uniform Hölder condition of order $\lambda - 2$ near $v = 0$. Consider the case

$$ds^2 = h(v)(du^2 + dv^2) \tag{6}$$

of (1). That (6) is the first fundamental form on a surface S of class C'' follows from the Remark [3], p. 138. In fact, such an S is given by $S\colon X = (f(v)\cos u, f(v)\sin u, g(v))$, where

$$f(v) = h^{\frac{1}{2}}(v) \text{ and } g(v) = \int_0^v (h - h_v{}^2/4h^2)^{\frac{1}{2}} dt$$

(the reality of f and g for small $| v |$ follows from (5)).

The differential equations of the geodesics, belonging to (6), are

$$(h(v)u')' = 0, \quad (h(v)v')' - \tfrac{1}{2}h_v(v)(u'^2 + v'^2) = 0, \tag{7}$$

where the prime denotes differentiation with respect to a variable which does not occur explicitly and which can be assumed to be the arc-length. In this case, one has the two first integrals

$$h(v)u' = \text{const.}, \quad h(v)(u'^2 + v'^2) = 1, \tag{8}$$

by (6).

3. Geodesic polar coordinates. In the case of the intial conditions (3), the first part of (5) shows that if $\phi \neq 0$ is fixed, the equations (8) become

$$(9) \qquad h(v)\,du/\cos\phi = dr, \qquad (\operatorname{sgn}\phi)h(v)\,(h(v) - \cos^2\phi)^{-\frac{1}{2}}\,dv = dr;$$

and that if $\phi = 0$, the equations (8) have the solution

$$(10) \qquad u = u(r,0) = r, \qquad v = v(r,0) = 0.$$

In order to avoid absolute value signs in what follows, it will be assumed that $\phi \geqq 0$, $r \geqq 0$, hence $v \geqq 0$. The definition (4) of h and the second part of (9) give

$$(11) \qquad \int_0^v (1 + t^\lambda)\,(t^\lambda + \sin^2\phi)^{-\frac{1}{2}}\,dt = r, \text{ where } v = v(r,\phi).$$

It is clear from (10) that $v(r,\phi) = o(1)$, as $\phi \to 0$, uniformly in r, say for $0 \leqq r \leqq \text{const.}$ Thus (11) implies that there exists a number $\theta = \theta(r,\phi)$, where $0 < \theta < 1$, satisfying $v(1 + o(1))\,(\theta v^\lambda + \sin^2\phi)^{-\frac{1}{2}} = r$. This relation can be written as $v^2(1 + o(1) - \theta r^2 v^{\lambda-2}) = r^2\sin^2\phi$. Hence, uniformly in r,

$$(12) \qquad v = r(1 + o(1))\sin\phi, \text{ as } \phi \to 0.$$

If one introduces $t/\sin\phi$ as a new integration variable, (11) becomes

$$\int_0^{v/\sin\phi} (1 + t^\lambda\sin^\lambda\phi)\,(1 + t^\lambda\sin^{\lambda-2}\phi)^{-\frac{1}{2}}\,dt = r.$$

By (12), the second factor of the integrand can be written as

$$1 - \tfrac{1}{2}t^\lambda(1 + o(1))\sin^{\lambda-2}\phi;$$

hence,

$$\int_0^{v/\sin\phi} (1 - \tfrac{1}{2}t^\lambda(1 + o(1))\sin^{\lambda-2}\phi)\,dt = r.$$

Thus $v/\sin\phi - \frac{1}{2}(\lambda + 1)^{-1}v^{1+\lambda}(1 + o(1))/\sin^3\phi = r$ or, by (12),

$$v = r\sin\phi + \tfrac{1}{2}(\lambda + 1)^{-1}r^{1+\lambda}(1 + o(1))\sin^{\lambda-1}\phi.$$

Since $\lambda - 1 < 2$ and $\sin\phi = \phi + O(\phi^2)$,

$$v(r,\phi) = r\sin\phi + \tfrac{1}{2}(1 + \lambda)^{-1}r^{1+\lambda}\phi^{\lambda-1}(1 + o(1)), \text{ as } \phi \to 0,$$

uniformly in r. The last formula line shows that $v_\phi(r,0) = r$, but since $\lambda - 1 < 2$, the partial derivative $v_{\phi\phi}(r,0)$ fails to exist for $r > 0$. In other words, the transformation $(u,v) \to (r,\phi)$ is not of class C''.

4. Geodesic parallel coordinates. Let ϕ be fixed and choose as the base geodesic K_ϕ: $u = u(r) = u(r, \phi)$, $v = v(r) = v(r, \phi)$, where K_ϕ is defined by (9) or (10) according as $0 < \phi < \pi$ or $\phi = 0$. Let σ denote arc-length along the geodesic $u = U(\sigma, \tau)$, $v = V(\sigma, \tau)$ which, for $\sigma = 0$, is orthogonal to K_ϕ at the point $r = \tau$. Thus σ, τ are geodesic parallel coordinates.

It can be supposed that U, V satisfy the respective initial conditions $U(0, \tau) = u(\tau)$, $V(0, \tau) = v(\tau)$ and $U_\sigma(0, \tau) = -v_r(\tau)$, $V_\sigma(0, \tau) = u_r(\tau)$. Then (8) becomes the system of differential equations

$$dU/d\sigma = -h(v(\tau))v_r(\tau)/h(V), \quad dV/d\sigma = (h(V) - h^2(v(\tau))v_r^2(\tau))^{\frac{1}{2}}/h(V).$$

Since $V(0, 0) = v(0) = 0$ and $v_r(0) = \sin \phi$, it follows that, at $(\sigma, \tau) = (0, 0)$, the square-root occurring on the right-hand side of the last differential equation is $\cos \phi$, which is not 0 if $\phi \neq \frac{1}{2}\pi$. Thus, if $\phi \neq \frac{1}{2}\pi$, the functions of U, V, τ on the right-hand sides of the differential equations are of class C'' for small $|U|$, $|V|$, $|\tau|$, since $v_r(\tau)$ is of class C'' by (9) or (10). Consequently, $V(\sigma, \tau)$, $V(\sigma, \tau)$ are of class C'' for small $|\sigma|$, $|\tau|$. In other words, the transformation $(u, v) \rightarrow (\sigma, \tau)$ to geodesic parallel coordinates is of class C'' when the base geodesic is K_ϕ, where $\phi \neq \frac{1}{2}\pi$.

On the other hand, it will be shown that if the geodesic $K_{\frac{1}{2}\pi}$,

$$u = 0, \quad v = v(r), \text{ where } \int_0^v h^{\frac{1}{2}}(t)\,dt = r, \tag{13}$$

is chosen as the base geodesic, the resulting transformation $(u, v) \rightarrow (\sigma, \tau)$ to geodesic parallel coordinates is not of class C''.

Let $u = U(\sigma, \tau)$, $v = V(\sigma, \tau)$ be the geodesic orthogonal to (13) at the point $r = \tau$, when the arc-length parameter σ is 0. As noted in [1], the methods there imply that the functions $U(\sigma, \tau)$, $V(\sigma, \tau)$ are of class C' on some set $|\sigma| \leqq$ const., $|\tau| \leqq$ Const. The assertion to be proved is that they are not of class C''.

In what follows, the letter v denotes $v(\tau)$, the function in (13) at $r = \tau$; the letter V, the function $V(\sigma, \tau)$. In order to avoid absolute value signs, it will be supposed that $\sigma \geqq 0$, $\tau \geqq 0$, hence $v \geqq 0$, $V \geqq 0$. All square roots will be non-negative.

The functions U, V satisfy the respective initial condition $U(0, \tau) = 0$, $V(0, \tau) = v(\tau)$ and $U_\sigma(0, \tau) = h^{-\frac{1}{2}}(v)$, $V_\sigma(0, \tau) = 0$. Thus, if $\tau \neq 0$ is fixed, (8) becomes

$$h(V)\,dU = h^{\frac{1}{2}}(v)\,d\sigma, \qquad h(V)(h(V) - h(v))^{-\frac{1}{2}}\,dV = d\sigma; \tag{14}$$

while if $\tau = 0$, (8) has the solution

(15) $$U(\sigma, 0) = \sigma, \qquad V(\sigma, 0) = 0.$$

Consequently, (4) and an integration of the second relation in (14) give

(16) $$\int_v^V (1 + t^\lambda)(t^\lambda - v^\lambda)^{-\frac{1}{2}}\, dt = \sigma.$$

If $(t^\lambda - v^\lambda)^{\frac{1}{2}}$ is introduced as a new integration variable, then (16) becomes

(17) $$\int_0^L (1 + t^2 + v^\lambda)(t^2 + v^\lambda)^{-1+1/\lambda}\, dt = \tfrac{1}{2}\lambda\sigma,$$

where

(18) $$L = (V^\lambda - v^\lambda)^{\frac{1}{2}}.$$

The second relation in (15) implies that

(19) $$V(\sigma, \tau) = \tau(V_\tau(\sigma, 0) + o(1)), \text{ as } \tau \to 0,$$

uniformly in σ, where $0 \leqq \sigma \leqq$ const.; in particular, if $\sigma = 0$, then by (13) and (4),

(20) $$v(\tau) = \tau(1 + o(1)), \text{ where } V_\tau(0, 0) = 1.$$

As t varies from 0 to L, the expression $t^2 + v^\lambda$ in (17) varies from $v^\lambda = V^\lambda(0, \tau)$ to $V^\lambda(\sigma, \tau)$. Thus there exists a $\theta_1 = \theta_1(\sigma, \tau)$, satisfying $0 < \theta_1 < 1$ and

$$L(1 + o(1))V^{1-\lambda}(\theta_1\sigma, \tau) = \tfrac{1}{2}\lambda\sigma.$$

By (19), the last relation is equivalent to

$$L = \tfrac{1}{2}\lambda\sigma\tau^{\lambda-1}(V_\tau{}^{\lambda-1}(\theta_1\sigma, 0) + o(1)).$$

The definition (18) of L and the mean-value theorem of differential calculus give $L^2 = \lambda(V - \sigma)m^{\lambda-1}$ for some number m satisfying $v < m < V$. In view of (19), this m is of the form $\tau(V_\tau(\theta_2\sigma, 0) + o(1))$, where $\theta_2 = \theta_2(\sigma, \tau)$ satisfies $0 < \theta_2 < 1$. Consequently, by the last formula line, $V(\sigma, \tau)$ is identical with

$$v(\tau) + (\lambda/4)\sigma^2\tau^{\lambda-1}V_\tau{}^{2\lambda-2}(\theta_1\sigma, 0)V_\tau{}^{1-\lambda}(\theta_2\sigma, 0)(1 + o(1)).$$

If this is divided by τ, and $\tau \to 0$, it is seen that $V_\tau(\sigma, 0) \equiv 1$. This is a consequences of (20) and the fact that $\lambda - 1 > 1$, while $V_\tau(\sigma, 0)$ keeps away from 0 for all sufficiently small $|\sigma|$. Thus the representation of $V(\sigma,\tau)$ by the last formula line simplifies to

$$V(\sigma,\tau)=v(\tau)+(\lambda/4)\sigma^2\tau^{\lambda-1}(1+o(1)).$$

Since $\lambda-1<2$ and $v(\tau)=\tau+O(\tau^2)$, it follows that

$$V(\sigma,\tau)=\tau+(\lambda/4)\sigma^2\tau^{\lambda-1}(1+o(1)),\ \text{as}\ \tau\to 0.$$

This equation and the inequality $\lambda-1<2$ make is clear that $V_{\tau\tau}$ fails to exist at any point $(\sigma, 0)$ where $\sigma\neq 0$. This completes the proof.

PARIS, FRANCE.

REFERENCES.

[1] P. Hartman, "On the local uniqueness of geodesics," *American Journal of Mathematics,* vol. 72 (1950), pp. 723-730.

[2] ——— and A. Wintner, "On the classical embedding problem in differential geometry," *ibid.*, pp. 553-564.

[3] ——— and A. Wintner, "On the problems of geodesics in the small," *ibid.*, vol. 73 (1951), pp. 132-148.

[4] ——— and A. Wintner, "On the asymptotic curves of a surface," *ibid.*, pp. 149-172.

ON LINEAR SECOND ORDER DIFFERENTIAL EQUATIONS WITH SMALL COEFFICIENTS.*

By PHILIP HARTMAN.**

Let $f(t)$ be a (real-valued) continuous function for large t-values, say for $T \leqq t < \infty$. The differential equation

$$x'' + fx = 0 \tag{1}$$

is said to be oscillatory or non-oscillatory according as one (hence every) solution $x = x(t) \not\equiv 0$ of (1) has or does not have an infinity of zeros on $T \leqq t < \infty$. Since

$$x'' + cx/t^2 = 0 \tag{2}$$

is oscillatory or non-oscillatory on $1 \leqq t < \infty$ according as $c > \frac{1}{4}$ or $c \leqq \frac{1}{4}$ (Kneser), the condition

$$F(t) = \lim_{X \to \infty} \int_t^X f(s)\,ds = \int_t^\infty f(s)\,ds \text{ is convergent} \tag{3}$$

is compatible with (1) being either oscillatory or non-oscillatory.

The following result was proved in [2], § 6:

(*) If (1) is non-oscillatory and (3) holds, then, for any solution $x = x(t) \not\equiv 0$ of (1), the logarithmic derivative $l = x'/x$ (which exists for large t-values) satisfies the Riccati differential equation

$$l' + l^2 + f = 0 \tag{4}$$

and the relations

$$\int^\infty l^2 dt < \infty \tag{5}$$

and

$$l(t) = F(t) + L(t), \tag{6}$$

---

* Received February 3, 1951.

** John Simon Guggenheim Memorial Foundation Fellow, on leave of absence from The Johns Hopkins University.

where $F(t)$ is given by (3), and $L(t)$ by

$$(7) \qquad L(t) = \int_t^\infty l^2(s)\,ds.$$

With the aid of (*), it was shown in [2], § 7, that if (3) holds and

$$(8) \qquad \int_T^\infty \exp(-A \int_T^t F(s)\,ds)\,dt < \infty,$$

where $A = 2$, then (1) is oscillatory. It was pointed out that, in view of the cases $c > \frac{1}{4}$ of (2), it is not possible to replace $A = 2$ by any number A exceeding 4, and the question was raised as to whether or not $A = 2$ can be replaced by $A = 4$. It will be shown in (I) below that the answer to this question is in the affirmative.

The criterion (8) for an oscillatory equation (1) is complementary to the following criterion of [2], § 2, for a non-oscillatory equation (1) (or even for the non-existence of any solution $x = x(t) \not\equiv 0$ with two zeros on $T \leqq t < \infty$):

$$(9) \qquad -\tfrac{3}{4} \leqq tF(t) \leqq \tfrac{1}{4} \text{ for } t \geqq T, \qquad (T > 0),$$

where (3) is assumed. If, in the proof given in [2], § 3, for the non-oscillatory character of (1) in the case (9), the function $y = F(t) + 1/(4t)$ is replaced by $y = F(t) + \gamma^2/t$, where γ is a fixed number in the range $\frac{1}{2} \leqq \gamma \leqq 1$, it is seen that (9) can be replaced by

$$(10) \qquad -\gamma - \gamma^2 \leqq tF(t) \leqq \gamma - \gamma^2 \text{ for } t \geqq T.$$

The criteria (10), belonging to different values of γ, are all distinct. But in a certain sense, the case $\gamma = \frac{1}{2}$ of (9) is "special." First, the $\frac{1}{4}$ on the right-hand side of (9) is the maximum value of $\gamma - \gamma^2$ for $\frac{1}{2} \leqq \gamma \leqq 1$ and cannot be replaced by a larger number. in view of the examples (2). On the other hand, as will be shown in (III) below, (9) implies that no solution of (1) satisfies

$$(11) \qquad 0 < \int_T^\infty x^2(t)\,dt < \infty,$$

and, in this connection, the $-\frac{3}{4}$ on the left-hand side of (9) cannot be replaced by a smaller number, in view of the examples (2). This statement has applications in the study of self-adjoint boundary problems on

$L_2(T,\infty)$ associated with (1). (It will be seen that the proof of (III) below can be modified to show that (10) implies that no solution of (1) satisfies

$$0 < \int_T^\infty x^{1/\gamma}(t)\,dt < \infty, \tag{11 bis}$$

and, in this connection, the number $-\gamma-\gamma^2$ on the left-hand side of (10) cannot be replaced by a smaller number; cf. the Remark following the proof of (III)).

With the aid of (*), it was shown in [2], § 8, that if (3) holds and (1) is non-oscillatory, then

$$\int_T^\infty \exp(B\int_T^t F(s)\,ds)\,dt = \infty, \tag{12}$$

where $B=2$, implies that no solution of (1) satisfies (11). It was pointed out that, in view of the cases $c < -\frac{3}{4}$ of (1), it is not possible to replace B by a number less than $\frac{4}{3}$, and the question was raised as to whether or not $B=2$ can be replaced by $B=\frac{4}{3}$; cf. also p. 370 of [2]. It will be shown in (III) below that the answer to this question is in the affirmative. (In particular, this implies the remark of the preceding paragraph concerning (11).)

(I) *Let* $f(t)$, *where* $T \leqq t < \infty$, *be a continuous function satisfying* (3). *Then* (8), *where* $A=4$, *implies that* (1) *is oscillatory. This assertion is false if* $A > 4$.

It may be remarked that if (8) holds for some A, where $0 < A < 4$, then it holds for $A=4$. In order to see this, let $G(t)=\exp(-\int_T^t F(s)ds)$. Then (8) means that the A-th power, $G^A(t)$, is of class $L_1(T,\infty)$. It follows from the definition of G that $G'/G=-F=o(1)$, as $t\to\infty$. Hence, $(G^A)'=AG^{A-1}G'=AG^A(G'/G)$ is $AG^Ao(1)$; so that $(G^A)'$ is of class $L_1(T,\infty)$. Consequently $\lim G^A(t)$, as $t\to\infty$, exists and is 0; that is, $G(t)=o(1)$. Finally, $G^4=G^AG^{4-A}=G^Ao(1)$ is of class $L_1(T,\infty)$, which proves the remark.

Proof of (I). Suppose that the positive assertion in (I) is false. Then any solution $x=x(t)\not\equiv 0$ of (1) has only a finite number of zeros. It can be supposed that T is chosen so large that a given solution $x = x(t)$ does not vanish for $t \geqq T$ (and that $T > 1$); the change of the lower limits of integration on the left-hand side of (8) clearly does not affect (8). It can also be supposed that $x(t) > 0$ for $t \geqq T$ (for otherwise x can be replaced by $-x$). Thus, if $l = x'/x$, then (4), (5) and (6) follow from (*).

The relations (4) and (6) imply that $L(t)$ satisfies the differential equation

$$L' + F^2 + 2FL + L^2 = 0; \tag{13}$$

and so,

$$L' + 4FL \leqq 0, \text{ that is, } (L(t) \exp (4 \int_T^t F(s)\,ds))' \leqq 0 \tag{14}$$

for $t \geqq T$. Since $L(t) \geqq 0$, by (7), it follows that

$$0 \leqq L(t) \leqq L(T) \exp (-4 \int_T^t F(s)\,ds) \text{ for } t \geqq T. \tag{15}$$

It follows from the assumptions of (I) that $\int^\infty L(t)\,dt < \infty$. The definition (7) of $L(t)$ shows that this is equivalent to

$$\int_T^\infty s l^2(s)\,ds < \infty. \tag{16}$$

Since $l(t)$ is the logarithmic derivative of $x(t)$,

$$\log(x(t)/x(T)) = \int_T^t l(s)\,ds. \tag{17}$$

If $l(s)$ is written as the product of $s^{\frac{1}{2}} l(s)$ and $s^{-\frac{1}{2}}$, Schwarz's inequality shows that

$$(\int_T^t l(s)\,ds)^2 \leqq (\int_T^t s l^2(s)\,ds)(\int_T^t s^{-1}\,ds). \tag{18}$$

Thus, by (16) and (17), $\log\ (x(t)/x(T)) \leqq C(\log t)^{\frac{1}{2}}$ for $t \geqq T (> 1)$ if C^2 denotes the value of the integral in (16). Hence $x(t) \leqq x(T) \exp C(\log t)^{\frac{1}{2}}$, which implies

$$\int_T^\infty x^{-2}(t)\,dt \geqq x^{-2}(T) \int_T^\infty \exp (-2C(\log t)^{\frac{1}{2}})\,dt = \infty.$$

Since $x = x(t) \not\equiv 0$ is an arbitrary (non-trivial) solution of (1), this leads to a contradiction. For, according to [1], p. 703, every non-oscillatory equation (1) has a solution $x = x(t) \not\equiv 0$ satisfying $\int^{\infty} x^{-2}(t)\,dt < \infty$. This proves the positive assertion in (I).

Since the negative assertion follows from the examples (2), the proof of (I) is complete.

The inequality (14), deduced from (13), has a dual, namely,

$$L' + 2(F^2 + L^2) \geqq 0. \tag{19}$$

It will be shown that this inequality and the main theorem of [2], § 1, lead to the following criterion that (1) be non-oscillatory:

(II) *Let* $f(t)$, *where* $T \leqq t < \infty$, *be a continuous function satisfying* (3). *If the differential equation*

$$z'' + 4F^2 z = 0 \tag{20}$$

is non-oscillatory, then (1) *is.*

It may be remarked that (II) is not a consequence of the Sturm comparison theorem. For example, if $f(t) = c/t^2$, as in (2), then $4F^2 = 4c^2/t^2$. Thus (20) is non-oscillatory if and only if $|c| \leqq \frac{1}{4}$, but $f(t) \leqq 4F^2$ if and only if either $c \leqq 0$ or $c \geqq \frac{1}{4}$.

Proof of (II). It can be supposed that T is chosen so large that (20) has a solution $z = z(t) > 0$ for $T \leqq t < \infty$. If $h = z'/z$ is the logarithmic derivative of z, then h satisfies the Riccati differential equation $h' + h^2 + 4F^2 = 0$; cf. (4). Let $g = \frac{1}{2}h$. Then $g' + 2(g^2 + F^2) = 0$; cf. (19).

Define $y = y(t)$ for $t \geqq T$ by placing $y = F(t) + g(t)$. Then $y(t)$ has a continuous derivative and satisfies the differential inequality $y' + y^2 + f = -(F - g)^2 \leqq 0$. According to [2], § 1, the existence of such a function $y = y(t)$ implies that a solution $x = x(t) \not\equiv 0$ cannot vanish twice for $T \leqq t < \infty$. This proves (II).

(III) *Let* $f(t)$, *where* $T \leqq t < \infty$, *be a continuous function satisfying* (3) *and, in addition, let* (1) *be non-oscillatory. Then, if* (12) *holds for* $B = \frac{4}{3}$, *no solution* $x = x(t)$ *of* (1) *satisfies* (11). *This assertion becomes false if* $B = \frac{4}{3}$ *is replaced by any* $B < \frac{4}{3}$.

If (12) holds for some $B > \frac{4}{3}$, then it holds for $B = \frac{4}{3}$. This can be verified as in the remark following the statement of (I).

Proof of (III). Suppose that the first assertion in (III) is false. Then (1) possesses a solution $x = x(t) \not\equiv 0$ which has only a finite number of zeros and satisfies (11). It can be supposed that the origin on the t-axis has been so chosen that $x(t) \neq 0$ for $0 \leqq t < \infty$, and that $x(t)$ has been normalized by

$$(21) \qquad x(0) = 1, \text{ so that } x(t) > 0 \text{ for } 0 \leqq t < \infty.$$

If $l = x'/x$, a quadrature of (6) gives

$$(22) \qquad \log x(t) = \int_0^t F(s)\,ds + \int_0^t L(s)\,ds.$$

But $\int_0^t L(s)\,ds = \int_0^t sl^2(s)\,ds + t\int_t^\infty l^2(s)\,ds$, by (7); hence

$$(23) \qquad \int_0^t L(s)\,ds \geqq \int_1^t sl^2(s)\,ds \text{ for } t \geqq 1.$$

By (17) and the case $T = 1$ of (18),

$$(24) \qquad \int_1^t sl^2(s)\,ds \geqq (\log mx(t))^2/\log t, \text{ where } m = 1/x(1) > 0.$$

Hence, by (22) and (23),

$$(25) \qquad \int_0^t F(s)\,ds \leqq \log x(t) - (\log mx(t))^2/\log t \text{ for } t > 1.$$

Since, by assumption, $x = x(t)$ is of class $L_2(0,\infty)$ and since (5) holds, the product $(x'/x)x = x'$ is of class $L_1(0,\infty)$. Hence $\lim x(t)$, as $t \to \infty$, exists and is 0. In particular, $\log mx(t) < 0$ for large t. Consequently, (24) implies that, for large t,

$$\log mx(t) \geqq -(\log t)^{\frac{1}{2}}\Big(\int_1^t sl^2(s)\,ds\Big)^{\frac{1}{2}},$$

and so

$$\infty > m^2\int^\infty x^2(t)\,dt \geqq \int^\infty \exp\Big\{-2(\log t)^{\frac{1}{2}}\Big(\int_1^t sl^2(s)\,ds\Big)^{\frac{1}{2}}\Big\}\,dt.$$

The integrand of the last (convergent) integral is monotone decreasing and therefore is $o(1/t)$, as $t \to \infty$ (Abel). Thus, as $t \to \infty$,

$$-2(\log t)^{\frac{1}{2}}\Big(\int_1^t sl^2(s)\,ds\Big)^{\frac{1}{2}}+\log t\to-\infty;$$

in particular $\int_1^t sl^2(s)\,ds \geqq \tfrac{1}{4}\log t$ for sufficiently large t. Hence, by (22) and (23),

$$\int_0^t F(s)\,ds \leqq \log x(t)-\tfrac{1}{4}\log t \text{ for large } t. \tag{26}$$

Let $t_0(>1)$ be chosen so large that the inequalities (25), (26) hold for $t \geqq t_0$. Let the half-line $t_0 \leqq t < \infty$ be divided into two mutually exclusive sets, R and S, by the requirements that

$$mx(t) \leqq t^{-\frac{1}{2}} \text{ if } t \text{ is in } R \tag{27}$$

and

$$mx(t) > t^{-\frac{1}{2}} \text{ if } t \text{ is in } S. \tag{28}$$

In the first case, when (27) holds, $\log mx(t) \leqq -\tfrac{1}{2}\log t(<0)$, which implies that $(\log mx(t))^2/\log t \geqq -\tfrac{1}{2}\log mx(t)$. Thus, by (25),

$$\int_0^t F(s)\,ds \leqq \log x(t)+\tfrac{1}{2}\log mx(t)=\tfrac{3}{2}\log x(t)+\tfrac{1}{2}\log m$$

if t is in R. Consequently, an integration over the set R gives

$$\int_R \exp\Big(\tfrac{4}{3}\int_0^t F(s)\,ds\Big)dt \leqq m^{\frac{2}{3}}\int_R x^2(t)\,dt<\infty. \tag{29}$$

On the other hand, (26) gives

$$\int_S \exp\Big(\tfrac{4}{3}\int_0^t F(s)\,ds\Big)dt \leqq \int_S x^{\frac{4}{3}}(t)\,t^{-\frac{1}{3}}dt,$$

while the inequality of Hölder shows that

$$\Big(\int_S x^{\frac{4}{3}}(t)\,t^{-\frac{1}{3}}dt\Big)^3 \leqq \Big(\int_S x^2(t)\,dt\Big)^2\Big(\int_S t^{-1}dt\Big).$$

Finally, (28) implies that $\infty > m^2\int_S x^2dt \geqq \int_S t^{-1}dt$. Hence, by the last two formula lines,

$$(30) \qquad \int_S \exp(\tfrac{4}{3}\int_0^t F(s)\,ds)\,dt < \infty.$$

Since $R + S = [t_0, \infty)$, the assumption, that (12) holds when $B = \frac{4}{3}$, is contradicted by (29) and (30). This completes the proof of the first assertion in (III).

Since the second assertion is a consequence of the examples (2), the proof of (III) is complete.

Remark. The theorem (III) remains true if the number $\frac{4}{3}$ and the formula (11), in the statement of (III), are replaced by the number $(\gamma + \gamma^2)^{-1}$, where $\frac{1}{2} \leqq \gamma \leqq 1$, and the formula (11 bis), respectively.

In view of the examples (2), it is sufficient to consider the first assertion of the modified theorem (III). The derivation of (25) remains valid. A slight modification of the derivation of (26) leads to

$$(26\text{ bis}) \qquad \int_0^t F(s)\,ds \leqq \log x(t) - \gamma^2 \log t \text{ for large } t.$$

If $t^{-\frac{1}{2}}$, in (27) and (28), is replaced by $t^{-\gamma}$, the modified version of (III) follows from (25) and (26 bis), as (III) does from (25) and (26).

PARIS, FRANCE.

REFERENCES.

[1] P. Hartman, "Differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[2] A. Wintner, "On the non-existence of conjugate points," *American Journal of Mathematics*, vol. 73 (1951), pp. 368-380.

ERRATA.

P. Hartman, "The number of L^2-solutions of $x'' = q(t)x = 0$," this Journal, vol. 73, pp. 635-645.

Page 637, in the last integral of formula (10), read $dQ^{-2}(s)$ instead of $dQ^2(s)$.

I. I. Hirschman, Jr., "On the distribution of the zeros of functions belonging to certain quasi-analytic classes," this Journal, vol. 72, pp. 396-406.

On page 403, line 3 from bottom, the upper integration limit should read $\mathfrak{H}^*(t)$, instead of u.

On page 404, line 1, omit the factor $\pi/2$.

Aurel Wintner, "On the Hölder restrictions in the theory of partial differential equations," this Journal, vol. 72, pp. 731-738.

Page 733, formula (2), read $\Delta u + \{$ instead of $u + \{$.

Page 733, line 3 from bottom, read (b) instead of (4).